

Setting priorities is painful but necessary

The United States needs a credible and far-sighted body to provide the government with forthright advice on scientific priorities. Scientists should support the National Science Board in its efforts to fulfil this task.

Is the US National Science Board playing with fire? Some of its own members fear that it may be doing so, as the board seeks to extend its remit from the narrow supervision of the National Science Foundation — the principal task for which it was created shortly after the Second World War — to the broader role of helping the federal government to coordinate all of its research activities (see page 220).

But someone has to do it. The 'endless frontier' of scientific discovery continues to expand in such marvellous and unexpected ways that no country, not even the United States, can push all of its boundaries at once. The one thing that is clear from the murky deal on a balanced budget that has just been reached by President Bill Clinton and the US Congress is that discretionary spending, of which science and technology are a large component part, will contract in real terms over the next five years, bringing difficult decisions in its wake.

The current arrangement that passes for US science policy, under which the president's political priorities collide somewhat haphazardly each summer with those of a dozen or so leaders in the Congress, does not serve the country well. Such a process is ill-equipped to make the tough choices that will undoubtedly have to be made as the United States' dominance of the world of science passes its peak, and other countries, both in Europe and among the emerging economies of Asia, come to rival its excellence in various fields.

The National Science Board is uniquely well-placed to offer advice on these choices: its founding charter of 1950, after all, asked it to offer science policy advice to the president and to the Congress. For a variety

of reasons, the board has never really carried out this task.

Twenty years ago, the Congress was sufficiently frustrated by the resulting vacuum to pass laws establishing the Office of Science and Technology Policy (OSTP), to lead science policy from the White House. That has not always worked either: as a priority-setting body in successive administrations, OSTP tends to have been increasingly ignored in the White House and mistrusted in the Congress.

It should not be beyond the wit of a body such as the National Science Board to draw up a convincing science policy for the United States. Studies by the Committee of Science, Engineering and Public Policy at the National Academy of Sciences have already shown how priorities could be identified. A panel of the academy, chaired by its former president (and presidential science adviser), Frank Press, laid out guidelines two years ago on how they could be implemented.

The fragmentation of control over science policy in the Congress remains a major obstacle to any such implementation. But there are strong indications that specific policy recommendations from a credible source would be well-received on Capitol Hill.

It is hardly appropriate for scientists to complain incessantly about the arbitrariness of the congressional budget process while steadfastly refusing to offer much-needed advice on what should be funded and what should not. The National Science Board is at least attempting to break this unseemly silence. It will be a sad reflection on the true priorities of US scientists if the attempt is snuffed out by those who fear the outcome of an honest attempt at selecting research priorities. □

Now for the hard part

Britain's Labour party still needs to show how it intends to put science 'at the heart of government'.

Forget 100 days. Britain's new Labour government has already hit the ground running, as it had pledged. There have been some unexpected — and significant — announcements, such as the decision to give the Bank of England full independence to set interest rates. And there have also been welcome moves already promised, such as the creation of a separate food standards agency.

But where is science in all this? Those who had been pressing Labour to appoint a cabinet minister solely responsible for science were disappointed, although perhaps not surprised, as the party had previously declined to make this commitment. Equally unsurprising is that there has not been any move to remove responsibility for the science base from the Department of Trade and Industry.

Even if some disagree with these decisions, there are sound reasons for both, related to the need to establish strong links between science and the country's industrial base. But there remains a genuine concern over what little has yet emerged about how science will be handled at the top levels of government. In particular, while responsibility for science has been given to a full minister, the portfolio has been merged not only with manufacturing but also with energy — an area which, only a few years ago, merited a whole ministry on its own.

No-one disputes either the energy or the commitment of John Battle, the opposition spokesman on science between 1994 and 1995, who has been selected to fill this post. But with many pressing demands on his agenda, such as energy regulation, will he really have the time to devote to science that it both deserves and requires?

The Labour party now has a considerable pool of talent in its ranks. Some have a proven track record, such as Anne Campbell, chair of the Parliamentary and Scientific Committee, and other members of the former House of Commons Select Committee on Science and Technology. New recruits also bring valuable skills. These include Ian Gibson, dean of biological science at the University of East Anglia and now Member of Parliament for Norwich (North), and Phyllis Starkey, formerly a policy and assessment expert with the Biotechnology and Biological Sciences Research Council, who now represents Milton Keynes.

In addition, several prominent scientists would no doubt welcome an opportunity to become closely involved with the development of an effective long-term and cross-government strategy (see above). Few will forget Tony Blair's commitments to place science 'at the heart of government'. □



US panel to back longer cloning ban, ducks thornier questions

[WASHINGTON] A US presidential advisory commission on bioethics is expected to recommend the continuation of a ban on the use of federal funds "for cloning of human beings" when it reports at the end of this month. The ban was initially declared by President Bill Clinton in March.

Extending the ban, according to members of the National Bioethics Advisory Commission (NBAC), will give US policy-makers and their advisers time to address and consult on the complex web of scientific, political and ethical issues raised by cloning.

According to one member of the commission, David Cox, a geneticist at Stanford Medical School, "[h]ow much of a stick to use and what stick to use is a complicated issue". He says that preserving the existing ban will buy time for "a more thorough national discussion".

The commission appears unanimous in its opposition to any use of newly discovered cloning technology to produce living humans. But separately, commissioners, while conceding the value of human cellular-level research that could profit from the new

technology, appear unlikely to try to protect it.

This is because it would breach a two-year-old congressional prohibition of federal funding of human embryo research. "This is not the time to revisit [the human embryo research issue]," said Harold Shapiro, NBAC chairman and the president of Princeton University, at an NBAC meeting on 2 May.

The new Clinton ban on government funding "for cloning of human beings" was in response to the publication of results of experiments in which researchers at Edinburgh's Roslin Institute cloned a sheep from a mammary cell of an adult ewe (see *Nature* 385, 810-813; 1997). On hearing the news, Clinton also asked the commission to complete within 90 days "a thorough review of the legal and ethical issues" associated with the new cloning technology and to recommend "possible federal actions" to prevent its abuse.

With the 90-day deadline fast approaching, the 18-member commission appears to have reached a level of agreement. "The minimum is a continued moratorium on cloning of human beings for implantation," says Ezekiel Emanuel, an associate professor of

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Where it all began: the birth of Dolly (left) continues to stir international controversy.

medicine at the Dana-Farber Cancer Institute in Boston. The real issue, he says, is whether the recommendation "is going to be stronger than that or not — and how".

Some argue that backing the presidential ban is a risk-free move. For example, Arthur Caplan, director of the Center for Bioethics at the University of Pennsylvania Medical Center in Philadelphia, says the recommendation is "possible but meaningless", as "no one has any plans" to experiment with human cloning in the near future.

Indeed, most commissioners appear keen to pronounce on more than simply federally funded activities. Laurie Flynn, executive director of the National Alliance for the Mentally Ill, and an NBAC member, says "[t]he looming concern about the private sector" dictates that "we have to speak to that directly and clearly".

Beyond urging a continued federal funding ban, the commission is considering two main policy options. The first would be a legislative ban on all cloning for implantation. The second is a continued ban on federal funding for such work, coupled with exhortations to the private sector to abstain voluntarily from such work, and to state licensing boards and professional societies to issue pronouncements against it.

It is hoped that such pronouncements would deter private experimentation through establishing standards for malpractice lawsuits. But commission members concede that relying on a voluntary ban would present risks. The history of private fertility clinics has shown their operators to be "not a very trustworthy lot", says Emanuel.

Other NBAC members have argued that a ban imposed through federal legislation would have the advantage of clarity, comprehensive coverage of public and private sectors

Roslin patents come under the spotlight

[PARIS] Embarrassment levels were high last week at the Roslin Institute in Edinburgh, which cloned Dolly the sheep, after it emerged that patents on the technology filed by the institute would cover all 'animals', including humans. The institute has taken a prominent stand against any attempt to use the technique to clone humans.

The scope of the patent applications was pointed out in a statement issued on 9 May by Rural Advancement Foundation International (RAFI), an Ottawa-based organization. If granted, the patents would represent "implicit acceptance of human cloning", says Hope Shand, research director of RAFI.

The patent applications have been registered at the World Intellectual Property Organization in Geneva, Switzerland, and will now go to patent offices in Europe, the United States and elsewhere for examination —

the approval process will take several years. Shand says that RAFI's statement is intended as a pre-emptive strike aimed at creating public awareness. She asserts that patent offices otherwise risked "blindly accepting" the patents with their current scope.

Shand hopes that patent offices will now restrict the scope of such patents. She also expresses the widely held sentiment that decisions on such moral choices should not be left to patent offices, but should rather be the subject of international agreements.

Indeed, one outcome of the debate about human cloning has been to trigger a broad consensus on the need for such agreements in bioethics. Recommendations aimed at international regulation of cloning were expected from the International Bioethics

Committee of the United Nations Educational, Scientific and Cultural Organization this week.

Harry Griffins, assistant director of the Roslin Institute, says the inclusion of humans in the patents was inadvertent. "Our intention at that time was to cover everything from farm animals to lab animals," he says. "It wasn't the intention to include humans."

He also retracts earlier statements in which he asserted that the scope of the patents would allow the Roslin Institute to prevent the technology falling into the wrong hands. He claims that these were off-the-cuff remarks. His statement had been pounced on by RAFI, who said it would be wrong to entrust the "ethics and fate of human cloning" to a single institute, and intends to mount a challenge against the patents. **Declan Butler**

and uniformity. These, they say, would be valuable in a situation where state laws against embryo research and cloning are already proliferating into a confusing patchwork.

But federal legislation could present its own risks both to individual reproductive freedom and to scientists. According to R. Alta Charo, a professor of law at the University of Wisconsin, who chaired a policy working group within the commission, a law without a "sunset" provision mandating expiry after a prescribed length of time "is likely to remain in place permanently".

That would constrain individual reproductive freedom if and when cloning technology reached a stage where it was deemed safe and ethical as a human reproductive technology. Some say this will never happen, but other NBAC members are unconvinced that the ethical arguments against cloning humans have been sufficiently developed to justify an indefinite ban.

Others point out that a law could be detrimental to research. Legislation could "easily be drafted inadvertently to cover the kind of simple, molecular DNA cloning that we do every day in laboratories around the country," says Charo.

The use of the new technology to conduct cellular-level research—for instance, on cellular ageing and differentiation—is already effectively barred by an existing congressional ban on human embryo research. Says Cox: "You can't do the cell-based work with any kind of nuclear transplantation without a human oocyte, and having that go through several divisions, which basically makes it an embryo".

NBAC has indicated that it will not challenge the congressional ban, which was attached as a rider to 1996 and 1997 appropriations bills, by, for example, explicitly fencing off human cellular-level research for protection from a broader law or moratorium aimed at cloning for implantation.

That may dismay scientists. Most of the heads of scientific societies who responded to a letter circulated by NBAC are said to support cloning human adult or embryonic nuclei for research in developmental biology and cellular therapeutics.

But even if NBAC were to explicitly recommend protecting cellular research using cloning, some believe Congress would ignore it because many conservatives are so adamantly opposed to embryo research. "Congress has already indicated it will not tolerate [human] embryo research. It's not going to change no matter what NBAC says or does," says Caplan.

That, says George Annas, a professor of public health at Boston University, leaves developmental biologists and others wishing to use the new technology with a considerable challenge. "The trick is to come up with a way to try to do cloning research without human embryos," he says. **Meredith Wadman**

France is urged to loosen ban on embryo research

[PARIS] France's national bioethics committee last week recommended that the ban on research on human embryos should be loosened to allow the use of embryonic stem cells in fundamental and therapeutic research.

The recommendation, which breaks a taboo on embryo research in France that has its origins in the conservative attitude of the Roman Catholic church, has been prompted by growing awareness that differentiated cell lines derived from these precursor cells could have huge potential therapeutic benefits.

Under legislation passed in 1994, research on human embryos is allowed only if it does not harm the 'integrity' of the embryo—in practice, a complete ban on embryo research.

The committee's recommendation would allow a dispensation only for research on embryonic stem cells. But it nevertheless marks a significant watershed, as it is the first time an unequivocal consensus has been reached in France in favour of allowing any form of human embryo research.

Embryonic stem cells, which occur at the blastocyte stage, can give rise to all tissue types when injected into older embryos, even though the cells themselves are unable to develop into a multicellular embryo or fetus.

In principle, it may be possible to cultivate such cells *ex vivo*, and to manipulate culture conditions to produce large quantities of bone-marrow stem cells for treating leukaemia, for example, or cerebral dopaminergic cells for treating Parkinson's disease. Although this is not yet technically possible in humans, the cloning techniques that produced Dolly the sheep now offer a possible solution (see previous page).

The committee proposes that research on stem cells should be allowed only using surplus conserved embryos, where the parents have confirmed in writing that they no longer wish to use their embryos, and have decided to terminate their conservation. There are more than 300 such embryos in France.

It qualifies its recommendation with a panoply of safeguards, including a ban on the creation of embryos specifically for research, on trade in stem cells and on their use to modify the germ line or create human clones.

Only one member of the committee, Father Olivier de Dinencin, a Jesuit priest, opposed the decision on embryo research, arguing that there was a choice between allowing surplus embryos to die naturally and using them for research. Taking the latter option, he argued, amounted to "reducing the embryo to an object, a material, for research".

Axel Kahn, director of the INSERM Laboratory of Research on Genetics and Molecular Pathology at the Cochin Institute

of Molecular Genetics in Paris, who is also a member of the committee, agrees that there is "incontestably" a danger that the committee's recommendation might be the beginning of a slippery slope towards the utilitarian argument against any limits on what can be done with embryos, particularly early embryos.

But Kahn argues that the new recommendation is a wise compromise between the extreme positions of those who believe embryos should not be experimented on at all and those who believe that early embryos are not persons, and that anything goes.

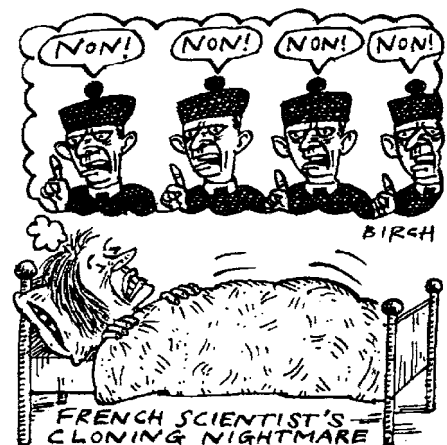
One interpretation of the committee's recommendation is that it confirms France's reluctance to rule on the question of whether the embryo becomes a person at the moment of conception or at some later point.

In Britain, for example, legislation passed in 1990 allows research on embryos up to fourteen days after fertilization. But there is a broad consensus in French society that the embryo is an 'evolutive entity' and that it is artificial to establish an arbitrary threshold before which everything is permitted and after which all research is banned.

If the committee's recommendations are adopted by the legislature during the scheduled revision of France's bioethics legislation in 1999, it would make French legislation in this area intermediate between that of Germany, where all human embryo research is banned, and that of both the United Kingdom and the United States.

The fate of the French ethics committee's recommendation will depend heavily on the outcome of next month's legislative elections. There is widespread hostility to human embryo research, especially among the conservative parties, and on the basis of the parliamentary debates preceding the introduction of bioethics legislation in 1994, many predict that a new conservative coalition may refuse to implement the panel's recommendations.

Declan Butler



Germany cuts funds for joint work with eastern Europe

[MUNICH] Germany's main research grant-giving body, the Deutsche Forschungsgemeinschaft (DFG), has decided to cut by a third its financial support for joint research projects between scientists from Germany and from central and eastern Europe.

The DFG's annual budget of more than DM1 billion (US\$581 million) comes from Germany's federal government. Although it is due to increase by 5 per cent over the next four years, the DM5-million now allocated to building scientific links with scientists from Russia and 20 other east and central European countries has been reduced by more than one third, to DM3.3 million.

The cuts take effect immediately. Cooperation with Russia will in particular suffer a significant setback, as the DFG had planned around 220 joint research projects with the Russian Foundation for Basic Research and the Russian Academy of Sciences this year, including bilateral scientific symposia.

"The cuts hit us at a time when our cooperation activities are becoming increasingly effective, as the criterion of scientific excellence has recently been taking precedence over political considerations in Russia," says Doris Schenk, the DFG official responsible for scientific collaboration with central and eastern Europe and the former Soviet Union.

Some Russian research institutes have become accessible to Western scientists only in the past four years. Before that, the top-level Russian researchers who worked in them — mostly theoretical physicists, mathematicians and chemists — were neither allowed to travel nor to publish their results.

"In many cases, we did not even know the names of these scientists," says Schenk. With DFG support they have been able to attend meetings and conferences in Germany to meet Western colleagues, often for the first time, and to become familiar with international scientific procedures.

But this process of *rapprochement* could now begin to decline. Germany's research minister, Jürgen Rüttgers, has announced a further cut in funds for the programme over the next three years. Schenk says: "This is a short-sighted policy since international collaborations with what used to be difficult partners need long-term continuity."

DFG officials are confident that they will be able to find money from within this year's budget to meet the contractual commitments for collaborative efforts to which they have already agreed. But they emphasize that this can only be done on a temporary basis by diverting the money from other activities, and that it does not represent a long-term solution.

Quirin Schlermeier

Japan calls in outside help to reform nuclear efforts

[TOKYO] Japan's beleaguered Science and Technology Agency (STA) is planning to call in foreign management consultants to carry out a complete review of the management of the Power Reactor and Nuclear Fuel Corporation (PNC). This unprecedented move follows a series of accidents at the corporation's nuclear facilities.

STA is under pressure to make radical changes at PNC, a semi-public organization under its jurisdiction with responsibility for a much of Japan's nuclear power research programme. PNC has been plagued with problems, leading to calls for it to be disbanded (see *Nature* 386, 746; 1997).

The STA is considering various foreign consultancy companies, including Arthur Andersen, one of the world's largest management consultants. According to STA officials, Japanese consultants are too "cosy" with the establishment to provide the independence needed of a third-party review.

The decision to review PNC's performance follows disclosures that PNC officials falsified reports about an explosion at a facility used for bitumenization of low-level radioactive waste at a nuclear fuel recycling plant in Tokai, northeast of Tokyo. The explosion exposed 37 PNC workers to low doses of radiation (see *Nature* 386, 209; 1997).

A preliminary report on the incident released by STA last week says the fire in drums used for storing nuclear waste was probably caused by a reaction between sodi-

um nitrate and sodium phosphate (in radioactive waste fluid) and asphalt used to solidify the nuclear waste. PNC had changed the procedure a few days before the accident, reducing the amount of asphalt used. According to Kaname Ikeda, director-general of STA's nuclear safety bureau, this may have led to "the troubles" that caused the fire.

The report states that a failure to extinguish the fire properly probably caused the explosion ten hours later. Ikeda says that there is a "defect in the management" which prevented the dissemination of knowledge and past experience to staff.

STA acknowledges for the first time in the report its own responsibility for the accident in not providing sufficient guidance to PNC. The agency has filed a criminal complaint against the organization and against PNC officials who tried to cover up false claims that the fire was immediately extinguished.

The consultants will be asked to focus on crisis management and notification systems in their review. Many analysts argue that Japan lacks sufficient expertise in crisis management. This was highlighted by the government's slow reaction to the Kobe earthquake in 1995, and the more recent disorganized response to an oil spill from a Russian tanker off the coast of Shimane Prefecture.

The consultants' report will be submitted to the recently formed PNC reform committee, headed by Horiyuki Yoshikawa, former president of Tokyo University. **Richard Nathan**

Synchrotron to move to science institute?

[TOKYO] The reorganization of Japan's Power Reactor and Nuclear Fuel Corporation (PNC) is likely to have far-reaching implications for other research organizations for which the Science and Technology Agency (STA) is responsible.

Serious consideration is being given to transferring some of PNC's responsibilities to the Japan Atomic Energy Research Institute (JAERI).

This might have a knock-on effect on the Institute of Physical and Chemical Research (RIKEN), which jointly runs Spring-8, the largest synchrotron in the world, with JAERI. In one scenario, Monju, Japan's controversial experimental

fast-breeder reactor transfers to a new organization and the Fugen advanced thermal reactor would go to JAERI. To stop over-expansion of JAERI, which already has 2,400 employees, RIKEN would take full responsibility for Spring-8.

PNC's uranium enrichment facilities in Ningyo-Toge, Okayama Prefecture, would be transferred to private electrical power companies in Japan or offered for sale to overseas companies in Britain and France.

Despite widespread reports of the proposals, STA officials say that none is under active consideration at this stage, and many other proposals have been put forward.

Nevertheless, one STA official says that the likelihood of the accelerator being handed over to RIKEN is "low". He argues that such a move would harm the atmosphere at RIKEN — which conducts multidisciplinary research — by concentrating too much research in one area. "You have to ask if it would be good for RIKEN and the future of nuclear power research in Japan," he says.

RIKEN has not commented officially on suggestions that it might take over Spring-8. But some senior institute officials believe RIKEN has the flexibility to take on the challenge. **Richard Nathan**

Science board is cautious on expanded policy role

[WASHINGTON] The National Science Board (NSB) has agreed to confer with other US research bodies about its plans to broaden its role and to help to coordinate research policy across the federal government.

Richard Zare, chair of the board and professor of chemistry at Stanford University in California, said last week that he was satisfied with a four-hour discussion of the plans by the board on 8 May. "This is moving in the direction I hoped it would," he said after the meeting. He intends to discuss the plans with the White House Office of Science and Technology Policy (OSTP) and other interested groups.

At its meeting, the board agreed to finalize a 'phase one' working paper proposing precise definitions for research and development (R&D), making the case for government spending on it, and suggesting "some kind of coordination" of R&D across the government.

After its approval, this paper will be released in August. According to Zare, the board will then discuss it with "other stakeholders" before producing more detailed proposals for how R&D priorities should be set and implemented across the government.

The US government says that it spends around \$70 billion each year on R&D. Of this, \$15 billion goes on basic research and \$25 billion on applied research and development. The remaining \$30 billion is spent on tasks such as the test and evaluation of new weapons by the Defense Department. Some experts, such as Frank Press, former president of the National Academy of Sciences, argue that such activities should not be classified as R&D.

Several board members warn of risks in the move to broaden the role of the NSB, whose main job is to oversee the National Science Foundation (NSF). They fear that

other research agencies will see it as an invasion of their turf.

Frank Rhodes, former president of Cornell University in New York state, and the previous chairman of the NSB, warned the board against trying to set priorities for the \$15 billion spent on basic research. The National Institutes of Health (NIH) account for \$6 billion of this, and NSF, the National Aeronautics and Space Administration and the Department of Energy about \$2 billion each. "We could be going into the ring with the NIH, and we've got to be very careful about that," he said. "It could be an uneven contest for us."

M. R. C. Greenwood, chancellor of the University of California at Santa Cruz and a former associate director of OSTP, made the same point: "There's a giant hole in our assumptions," she warned. "The largest part of this is health research, which has its own agency and its own advisory board."

Zare says that he has been in communication with Harold Varmus, the director of NIH, about the NSB's plans. Varmus is "interested", but cannot respond further until the NSB has some concrete proposals.

Ian Ross, former president of AT&T Bell Laboratories, who chairs a subpanel of the board which is preparing the working paper, said that it was going to be difficult to get consensus from the board on key issues, such as the breadth and depth of any priority-setting exercise, and who should conduct it.

Board members identified three questions that must be answered before such an exercise begins. First, would it be good for science? Second, would it be good for the NSB? And, as Shirley Malcom of the American Association for the Advancement of Science put it, "if we do set out priorities, who is going to listen to us?" **Colin Macilwain**

US academy staff face pay freeze after court ruling

[WASHINGTON] The salaries of 1,100 staff at the National Research Council (NRC), the executive arm of the US National Academy of Sciences, have been temporarily frozen as the council's management works out how to deal with recent court rulings requiring it to comply with the Federal Advisory Committee Act (FACA).

The freeze was implemented last month, before a final ruling by the US Court of Appeals last week that the NRC must, indeed, comply with FACA. This means the council's advisory committees must open their activities to public scrutiny. William Colglazier, executive director of the NRC, describes the freeze as "essentially a pause".

In a memorandum to staff, the presidents of the National Academies of Science and Engineering and the Institute of Medicine — the three main components of the NRC — said the freeze was needed because "we face some uncertainty in the short term about the academy's portfolio of contracts and work". Colglazier now says that the NRC's budget projections should allow the freeze to be lifted, but is unsure when.

Last week's court ruling was a serious, if not wholly unexpected, blow for the NRC (see *Nature* 386, 309; 1997). The judges ruled that the academy complex should be subject to FACA because a 1989 Supreme Court judgment, in a case not directly involving the academy, had said that it should.

The academy presidents will now confer with federal agencies, which sponsor most of the NRC's work. According to Colglazier, the complex is likely to choose between three approaches for each of its studies.

If the sponsor agrees, it will appoint a committee to conduct its work in the traditional way. In other studies, it will implement a new committee procedure designed to comply with FACA. But in some cases the NRC will appoint a 'principal investigator' to conduct a study, eliminating a committee and avoiding FACA jurisdiction. The third approach is likely to face legal challenge, Colglazier says.

The NRC will appeal against the Supreme Court judgment, hoping to win the support of the Department of Justice. If the Supreme Court hears the case it will do so later this year and issue a ruling next spring. But it only hears a small minority of the appeals brought to it. NRC officials are therefore also pursuing changes in the law that would exempt the academy complex from FACA. Congressional leaders are believed to be sympathetic, but the committees with jurisdiction are at present paralysed by partisan bickering about their investigations into President Bill Clinton's ethics. **Colin Macilwain**

Holland unveils telescope design entry

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS
UNAVAILABLE
FOR COPYRIGHT
REASONS

[LONDON] Researchers at the Netherlands Foundation for Research in Astronomy last week unveiled a design they hope will be chosen for the world's largest radiotelescope. SKAI, the Square Kilometer Array Interferometer (left), will connect a raft of smaller receivers spread out over 150 km.

The US\$150-million telescope is a collaboration between the Netherlands, Canada, Australia, India, China and the United States. An intergovernmental committee is scheduled to choose between rival designs in 2001.

Truce likely in battle over 'DNA-chip' patent rights

[PARIS] A leading British scientist and a US biotechnology company say they are keen to reach an amicable settlement over the European patent rights to so-called 'DNA-chips', a key genome technology, rather than face a protracted court battle in which either party might lose all.

The patent, which was granted in 1989, is held by Ed Southern of the University of Oxford, and is being challenged by Affymetrix of San Francisco. It covers the technique of using oligonucleotide arrays as a testing platform, and Southern has applied for a similar patent in the United States.

Southern invented the technique of Southern blotting, but never patented it; "I'm learning from my mistakes," he says. He is setting up a company, Oxford Gene Technology, to commercialize DNA-chip technology, particularly for disease genotyping and antisense applications.

But he says that he "feels quite strongly" that the company should adopt a "reasonable" licensing policy. "It's a broad-based core technology, and it would be wrong for one company to hold a monopoly position."

The DNA chip patent, however, is being challenged by Affymetrix, and independently by six other companies — including Hoffmann La Roche and Abbott Laboratories — in a formal opposition that may take several years to work its way through the European Patent Office.

The principle of the technology is similar to that of gel electrophoresis-based techniques of DNA hybridization. But it is much easier to automate and gives much higher throughputs. By combining light-directed synthetic techniques (photolithography) developed in the semiconductor industry with conventional oligonucleotide synthetic chemistry, it allows the automated production of up to 100,000 specific DNA sequences on the surface of microchips in a few hours.

Such chips are widely considered one of the most promising genome technologies. Using hybridization, the sequence of genes, and indeed whole genomes, can be quickly compared. This opens up possibilities for deducing the function of genes, and developing genetic tests and antisense therapies.

An indication of the interest in the technology comes from the creation last month of a consortium by Affymetrix, Bristol-Myers Squibb and Millennium Pharmaceuticals to support a \$40-million five-year research programme aimed at developing gene-based technologies to elucidate the function of genes. The work will be carried out at the Whitehead Institute/Massachusetts Institute of Technology Center for Genome Research.

Affymetrix, which is 34 per cent owned by Glaxo Wellcome, is widely acknowledged as the leader in the technology. It has already begun marketing its products in the United States, but its entry into Europe is being obstructed by Southern's patent.

In their challenge to the validity of the patent, Affymetrix and the other opponents allege that Southern's claims are not novel, not inventive, and that the specifications are insufficient to support the claims of the existing scope.

Southern disputes these claims. "I'm pretty confident that we will come through in a strong position," he says. "I think it was an original invention." Piers Pennant, a patent attorney at Stevens, Hewlett and Perkins, the London-based law firm representing the University of Oxford, says: "Affymetrix is the

main opponent — if we can answer their claims then all the others follow."

But he concedes that the opponents' claims concerning scope are "to some extent justified", adding that "the claims we were granted in Europe initially were broader than was justified". Southern has now filed amended claims of narrower scope, according to Pennant.

Conciliatory noises are being made by both parties. "I'd hope we can work this situation out without having to go through with the challenge," says Vern Norveil, a lawyer at Affymetrix. Southern agrees that both sides are willing to find common ground. "We are still talking [to Affymetrix and several of the other opponents]," he says.

In a case to be heard later this year in the Northern District Court of California, Affymetrix will face complaints that it has infringed two 1993 patents on DNA-chip technology held by another Californian company, Hyseq (see *Nature Biotechnology* 15, 406; 1997). Hyseq is also challenging Southern's patent.

Declan Butler

US farmers warm to bioengineered crops

[WASHINGTON] As genetically altered crops create increasing controversy in Europe, US farmers are reported to be planting roughly ten times more acres this spring with engineered corn and soya beans than they did when the products were introduced last year.

About 12 per cent of US soya bean fields are being planted with modified soya beans, and about 6 per cent of corn acreage with modified corn, according to estimates by the companies that developed the crops.

"It's here to stay, I think," says Ken Fawcett, a farmer in West Branch, Iowa, who sells the altered soya beans for Asgrow, a major seed company. Fawcett says that his sales of the Monsanto company's soya beans doubled this year.

Wayne Hoener, who oversees national sales of the soya beans for Asgrow, estimates that sales have tripled this year but that demand has increased "eight or ten times".

But opponents of the crops say that the companies' rosy figures will be short-lived. "Many farmers are unfortunately buying into the public relations and marketing hoopla of Monsanto and Novartis," says Jeremy Rifkin, president of the Foundation on Economic Trends, which last autumn launched a global boycott of the engineered corn and soya beans. "I think they're going to find that there are serious problems with the crops."

Jeff Wells, co-founder of Genetic ID, an Iowa company that tests crops for genetic alterations, says that the new numbers mask a wariness among many US farmers about the implications of increasing consumer resistance to genetically altered foods in Europe (see *Nature* 386, 532 & 745; 1997.)

"A significant segment of farmers don't want to be caught on the wrong side of the market," he says. As a result, "a lot" of farmers are keeping their options open by planting "a little" of the crops.

The soya beans are engineered to resist damage by Roundup, a herbicide made by Monsanto. The corn, whose lead maker is Novartis Seeds (formerly Ciba Seeds), has a gene from *Bacillus thuringiensis* (*Bt*), a soil bacterium. The gene codes for a protein toxic to the European corn borer, a common pest.

Monsanto claims that the increased use of its soya beans — last year they were planted in just under 2 per cent of US acreage — is a tribute to their effectiveness. The company says they allow farmers to avoid expensive pesticide regimens without losing yield. Monsanto's seeds sell for about \$24 a bag, compared to \$16 to \$18 for unmodified seeds.

Novartis Seeds says the tenfold increase in acreage of *Bt* corn planted this year is proof that the crop is winning over farmers. "If that is not a success story, then I don't know what is," says Albin Hubscher, Novartis vice-president for marketing. He says that the company has sold out of 15 of the 18 genetically modified hybrids it produced.

Novartis says its corn accounts for 70 to 80 per cent of the engineered corn planted this year. It costs \$124 a bag, compared to \$80 for unengineered corn. Some farmers do not think it is worth the price. Wallie Hardie of North Dakota, who planted 150 acres of *Bt* corn last year, is not planting any this year because yield was not appreciably different from unmodified corn.

Meredith Wadman

University cuts 'lack real justification'

[SYDNEY] Australian universities have appealed to the government to increase rather than cut funding of higher education to avoid the country falling behind its counterparts in South-East Asia — and losing its academic reputation in the process.

The appeal was made against a backdrop of increasing unrest on university campuses. It reflects growing concern about the impact of changes implemented abruptly last year by Amanda Vanstone, education minister in the then newly elected Coalition government of Prime Minister John Howard.

Vanstone has already hinted in a Senate committee hearing that the government may break an undertaking that cuts totalling 5 per cent over the next three years would not be extended beyond that period. Vanstone suggested that there might be a further 1 per cent cut in a fourth year.

At the time of last year's announcement, universities complained bitterly that cuts had been made to operating grants and student fees increased purely for financial reasons, with little justification in terms of policy (see *Nature* 382, 569; 1996).

After the cuts had been ratified by the

Senate, Vanstone set up early this year the first national review of the role of, demand for and provision of higher education for a decade. The seven-member review panel is chaired by Roderick West, the highly regarded former headmaster of a private boys' school, and includes three senior university officials.

The panel is due to report by the end of 1997. Its work coincides with growing uncertainty and unrest in universities and public research institutions. Staff have staged stoppages because of forced redundancies and pay claims, and students have occupied university offices and demonstrated noisily whenever Vanstone has appeared in public.

One of the most contentious changes has been to allow universities that wish to do so to charge full fees to Australian students — up to a maximum of 25 per cent of total undergraduates in any course — to make up



Gale: warns prestige may be undermined.

for the loss of income due to government cuts. Until now, only students from overseas have been charged full fees. These are a significant source of income for universities, producing an estimated A\$600 million (US\$470 million) in 1996.

The two oldest universities, Sydney and Melbourne, were among the first to announce plans to offer full-fee places, Sydney's ranging from A\$33,000 for a three-year degree in humanities to A\$100,000 for a four-year degree in dentistry.

But some universities, including Australian National University and LaTrobe University, have said that they will rely solely on the present system of partially deferred fees. They argue that they remain committed to a policy of wide access and equity.

Last weekend saw violent student demonstrations in both Melbourne and Sydney over the introduction of full fees for Australians. At Sydney, students broke into the vice-chancellor's office, from which they were evicted by the police.

Staff and students also demonstrated at the University of New South Wales last Friday (9 May) because of a decision by the university's council to award the vice-chancellor a pay rise of A\$50,000 at a time when the university is considering introducing full fees.

Another recent ruling that has shaken science-based faculties, already concerned about declining enrolments for first degrees, is that postgraduate awards are to be taxed. Last week, an alliance of staff and students vowed to continue its campaign against the many changes until the next election, due in less than two years.

The combined submission to the government on behalf of the country's 37 universities was issued by Fay Gale, president of the Australian Vice-Chancellors' Committee and vice-chancellor of the University of Western Australia. Gale said there was unanimous agreement that spending on higher education — including contributions from industry — should increase from 1.65 to 2 per cent of the country's gross domestic product by 2010.

Gale pointed out that Taiwan and Singapore now outstrip Australia in proportion of GDP spent on education at all levels by factors of two and three respectively. "In a relatively short time Australia's national influence, prestige and ultimately security could be undermined," she warned.

The vice-chancellors' committee has joined the Federation of Australian Scientific and Technological Societies in attacking the government's decision last year to cut the tax deduction for research in industry to 125 per cent. The vice-chancellors are seeking its restoration to 150 per cent, and the federation to 200 per cent.

Peter Pockley

Australian funding begins long drift downhill

[CANBERRA] The second budget of Australia's Coalition government, released in Canberra this week, provides no new initiatives for research, and is likely to generate concern among scientists about their prospects in a conservative political climate.

While the heavy cuts in last year's budget were justified in terms of a need to reduce a deficit inherited from the previous Labor government, Peter Costello, the Treasurer, said on Tuesday (13 May) that more cuts and revenue-raising measures were needed to place the budget in surplus in the financial year that starts on 1 July.

In his speech to parliament, Costello made no reference to science or technology. Nor did he refer to higher education (see above) or to any other form of intellectual or cultural activity — suggesting that the government does not see these as contributing to the

economic recovery it is predicting.

The amount of public funds to be committed to research and development next year show an increase over the current year, expressed in current dollars. But figures adjusted for inflation show a downward trend starting gently in 1997 and accelerating to 2001.

Science and technology minister Peter McGauran, who is responsible for one-fifth of the government's outlays on research spread over seven ministries, said in a statement issued with the budget that the government was "moving ahead with its strategy of making science spending more relevant and accountable".

McGauran repeated the priority for marine science that was announced last year. But he gave no indication of any new activities in this field — or in any of the government's research agencies.

Last month, McGauran

told the Cooperative Research Centres Association that their 66 centres, costing A\$142 million (US\$110 million) this year, are the "centrepiece" of government policy and that he sees them as "catalysts for economic growth" (see *Nature* 385, 191; 1997). In his budget statement, he argued that the government had "maintained its commitment to the program" and that funding was "fully in place".

But figures in the forward estimates of the budget reveal that funding for the centres will be cut by 5.5 per cent in 1998–99. The association has been told that total funding will be cut by A\$1 million next year, and by A\$10 million in the following year.

Health and medical research faces a similar fate. After a small increase next year, forward estimates nosedive by 14 per cent in each of the next two years. Defence research drops this year by 12 per cent.

P.P.

Hughes confirms its faith in excellence

Colin Macilwain

Over the past decade, the Howard Hughes Medical Institute has seen a steady escalation in the value of both its endowment and its reputation as a source of support for high-quality biomedical research.

[WASHINGTON] For a fortunate group of America's best life scientists, things will never be the same again after next Monday, 19 May. On that day, this élite will join the staff of the Howard Hughes Medical Institute (HHMI) as part of the philanthropic organization's largest expansion — their number comfortably exceeding the 44 investigators added in 1994.

The researchers will be assured of generous support from HHMI for five or seven years — and of virtually unfettered freedom to interrogate nature in whichever way they choose.

The emergence of Hughes over the past decade as a major player in the support of biomedical research means that many of the very best researchers are no longer dependent on public funds, a major departure for biomedical research in the United States.

HHMI will spend \$450 million this year on research, grants and programmes (see graph). In comparison, the publicly funded National Institutes of Health spend \$12.5 billion. But, according to the Institute for Scientific Information (ISI) in Philadelphia, Howard Hughes investigators have in recent years produced about a quarter of the most cited biomedical research papers.

In the first two months of this year, 12 of the 50 most frequently cited biomedical papers published since 1995 were written by HHMI individuals. "The trend is clearly for the private entities to fund the best work," says David Pendlebury, a senior analyst at ISI. "In a broad sense, this is a privatization."

Monday's announcement will further strengthen HHMI's existing core of 279 investigators, based at 60 medical colleges and research institutes across the United States.

HHMI was founded by Howard Hughes in 1953 as a small charity with large tax advantages. It came into its own in 1985, when the complex unravelling of the eccentric billionaire's estate combined with the sale of the Hughes Aircraft corporation for \$5 billion to leave the institute with the largest endowment of any research trust in the world.

Pressed by the Inland Revenue Service to spend at least 3.5 per cent of the endowment on medical research every year, HHMI quickly embarked on a programme to build and operate laboratories at selected universities and medical schools across the United States.

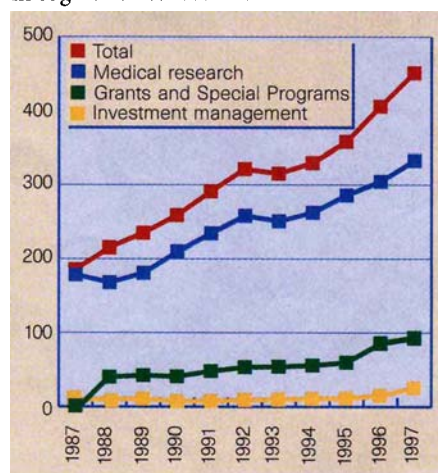
Under the leadership of Purnell Choppin,

a virologist and former administrator at Rockefeller University, New York, who joined HHMI as chief scientific officer in 1985 and became its president two years later, it has since moved steadily towards supporting individual investigators, wherever they work. All the new investigators — who will be listed at www.hhmi.org/whatsnew/ on the Internet on 19 May — will become HHMI employees, and their universities will enter agreements with Hughes to support their work at their existing laboratories.

The winners were chosen in March by the 45 members of HHMI's medical advisory and scientific review boards from 370 nominations received from 200 universities, medical schools and other research institutions.

Max Cowan, chief scientific officer at HHMI, says that he is aware of the risk that this process may merely perpetuate HHMI's existing scientific interests. "You're always biased towards the things you already know," he says. But Cowan promises that the new appointments will ensure that HHMI broadens its interests within each of the five disciplines around which it organizes its work — immunology, cell biology, genetics, neuroscience and structural biology.

In structural biology, HHMI will add specialists in electromicroscopy and nuclear magnetic resonance, and in bio-organic chemistry. In neuroscience, where HHMI began with a sharp focus on cellular work, the newcomers will include people working in cognitive neuroscience. There will also be



Spending by the Howard Hughes Medical Institute, 1987–97.

more researchers with strong clinical interests, Cowan says.

HHMI's charter restricts its interest to biomedical research — which is just as well, in Cowan's view. "One of our strengths is that we haven't tried to become a 'science institute'," he says. "You can think of plenty of philanthropic organizations who have tried to become all things to all people."

Aided by a strong stock market, Hughes' endowment has surged in value. Last year, as HHMI spent \$406 million, its net assets grew by \$446 million, to \$8.8 billion. Asked why it does not spend more, Choppin says that the research expenditure — \$304 million in 1996 — is determined by what it considers sustainable "in the very long term."

"We've had the luxury of expanding the endowment because the market has been favourable," says Choppin. HHMI has expanded its grant programmes, which support science education at various levels in the United States, as well as research in eastern Europe, Canada and South America. Unlike the research programme, Choppin says, these can be cut back quickly if the market deteriorates.

Choppin dismisses the suggestion that Hughes, by choosing to fund only a few, well-established scientists, discriminates against the young or adventurous. He says HHMI supports 350 PhD students — 100 medical students interested in research, and 700 post-doctoral associates.

According to David Blake, vice-president for research at the American Association of Medical Colleges and former dean of research at Johns Hopkins University in Baltimore, Maryland, many universities have experienced problems integrating HHMI investigators and their publicly funded colleagues.

Blake says Hopkins pioneered arrangements to ensure that facilities were shared and that each group had comparable salaries and benefits. "Some institutions didn't do that, and great animosity developed."

No one doubts the excellence of HHMI's work. But some critics of the government's system for supporting research are using it to highlight what they regard as the shortcomings of the National Institutes of Health. "Hughes is élitist, and so it should be," says Cecil Fox, who runs a private immunology laboratory in Gaithersburg, Maryland. "I can always tell an HHMI investigator — they have a class and a style that is being passed on. Publicly funded science will continue to be mediocre — and it'll continue to get worse."

But political support for publicly funded biomedical research remains strong, and seems unthreatened by the emergence of alternative funding sources such as HHMI. "Both models work," says Tony Fauci, director of the National Institute of Allergy and Infectious Diseases. "There are enough outstanding scientists for them to complement each other, rather than compete." □

Sweden urged to stay in CERN but to quit molecular biology lab

[MUNICH] Sweden should give up its membership of the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany, according to a government advisory committee, as part of a package to save SKr150 million (US\$19.5 million) in the government's research budget. But the committee does not recommend pulling out of the European Laboratory for Particle Physics (CERN) in Geneva, Switzerland, one of the options it had been considering (see *Nature* 386, 636; 1997).

Other proposals include withdrawing from the Vienna-based International Institute for Applied Systems Analysis, and reducing the national fusion research programme. In addition, national polar and space research, as well as some small research grant programmes, should receive less money, says the committee, as alternative funds are now available.

According to Torbjörn Fagerström, a professor of theoretical ecology at the University of Lund, and the committee's general secretary, EMBL was singled out as a means of making savings on European scientific activities because Sweden has a strong base in cellular and molecular

biology, and is therefore less reliant on international collaboration. The report is now being considered by Carl Tham, the minister of education and research who will make a proposal to parliament next month.

WHO and Unesco link hands on malaria

[PARIS] The World Health Organization (WHO) and the United Nations Educational, Scientific and Cultural Organization (Unesco) have agreed to cooperate in promoting education about malaria. The organizations believe that improved education — for example, on personal protection measures, and early diagnosis and treatment — could have an impact on the prevalence of malaria, which affects 300 million to 500 million people annually and kills 1.5 million to 2.7 million (see *Nature* 386, 535; 1997). The two bodies will cooperate in the development of educational materials and media packs, and the training of teachers.

Study will investigate Gulf War illness

[LONDON] Britain's new armed forces minister, John Reid, has promised to take urgent action on illnesses suffered by Gulf War veterans. In a policy statement issued on 11 May, Reid announced that the

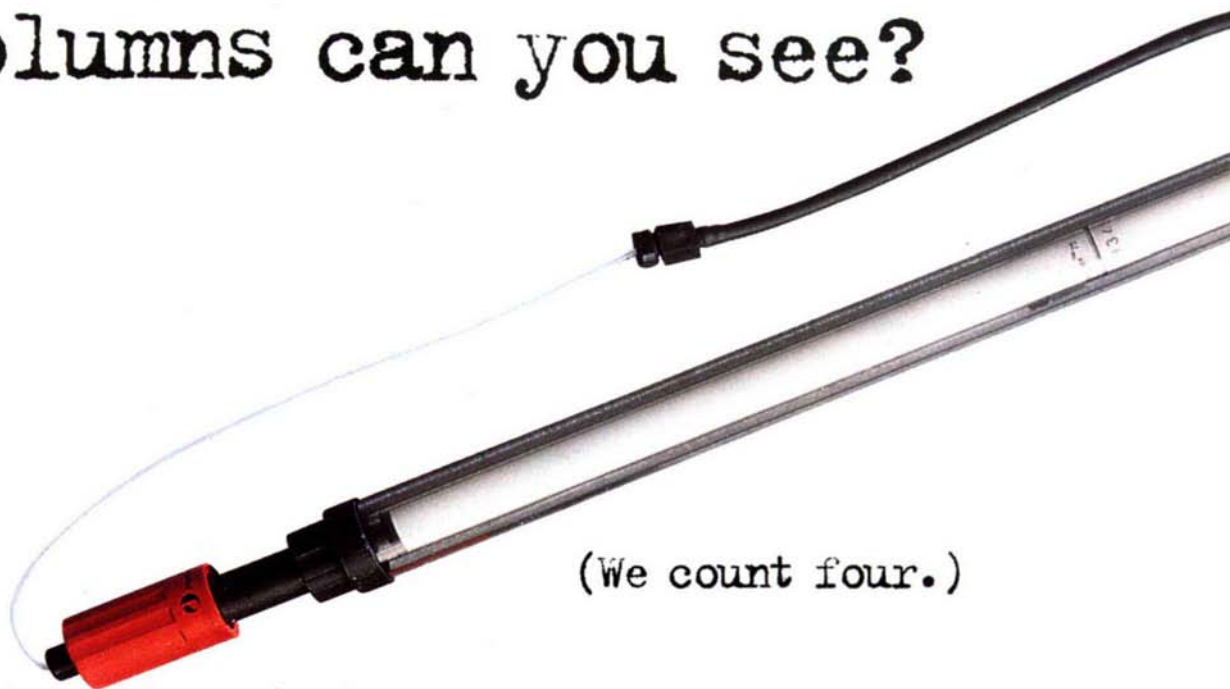
government is to launch a research project to investigate the causes of a range of illnesses collectively known as 'Gulf War syndrome'. This will be separate from the £3-million (US\$4.8 million) epidemiological study commissioned by the previous Conservative government last November on the advice of the Medical Research Council (see *Nature* 384, 604; 1996). It will investigate the effects of vaccines and anti-nerve gas drugs given to service personnel to counter biological and chemical weapons.

Brussels beefs up its food advisory process

[PARIS] The European Commission has adopted details of a comprehensive overhaul of its food policy and system of scientific advisory committees, prompted by the bovine spongiform encephalopathy crisis. The commission says that "excellence, independence and transparency" will now be the three principles guiding its gathering of scientific advice. To that end, it intends to recruit "eminent" scientists, who must have no possible conflicts of interest in their role as advisers.

Emma Bonino's directorate for consumer policy and health protection (DG24) is to be reinforced with 273 extra staff through redeployment of staff from other directorates, and recruitment. Her

How many
gel filtration
columns can you see?



(We count four.)

directorates now regroup all the commission's advisory committees, previously spread throughout the commission. Their work will be overseen by a new scientific steering committee.

Food standard agency meets Labour's promise

[LONDON] Tony Blair, the new British prime minister, has promised to press ahead with proposals to set up a food standards agency independent of the Ministry of Agriculture, Fisheries and Food, and reporting to the Department of Health. The proposals are outlined in a report by Philip James, director of the Rowett Research Institute at the University of Aberdeen.

The report was commissioned by Blair in March as part of his party's proposals to rectify the loss of public confidence in British food policy. The report says the agency should be managed by a 10-member commission dominated by "public and consumer interests". It also calls for government advisory committees to meet in public.

Ireland 'still lacks a credible science base'

[MUNICH] The Royal Irish Academy has criticized Ireland's first white paper (policy document) on science, technology and

innovation for failing to put forward a "coherent strategy for improving Ireland's research and development strategy and its climate for research". In its formal response to the white paper, the academy criticizes the document's failure to address the chronic underfunding of basic research, saying that Ireland is left "well short of what it needs for a credible national research programme". The importance of basic research had been stressed in the 1995 report of the government's Science, Technology and Innovation Advisory Council.

Russian science wins concessions on tax

[MOSCOW] Russia's science ministry has emerged victorious after a bruising encounter with the finance ministry over plans to levy tax on grants to research institutions. Under the new tax code, research grants are to remain tax-exempt.

Some Russian scientists remain displeased, however, as tax-exempt status will be refused to institutions that get more than 30 per cent of their funds from non-government sources. "The new tax code [is] the last nail in the Russian science coffin," says Gennady Mesyats, vice-president of the Russian Academy of Sciences.

Mesyats is also concerned that a new property tax, and a 20 per cent indirect tax

on sales of research products or services, will penalize the more successful institutions. The new tax code has been presented to the Russian parliament. If signed into law, it will remain in force until January 1999.

Placebo use defended in Third World studies

[WASHINGTON] US biomedical leaders last week defended the use of placebos in trials of the anti-AIDS drug AZT in pregnant women in Africa and Asia. Harold Varmus, the director of the National Institutes of Health, and David Satcher, the director of the Centers for Disease Control and Prevention, told a House of Representatives subcommittee that the trials, in countries including the Ivory Coast and Thailand, are scientifically and ethically justified.

Christopher Shays, the chairman of the subcommittee on human resources of the House Government Reform and Oversight Committee, challenged trials with placebos overseas when long-course AZT has been proved effective in reducing perinatal AIDS transmission in developed countries. Satcher said that "the international community has never accepted the [long-course regimen] as appropriate for developing countries", and that studies with placebos were necessary to determine the value of short-course AZT in those countries.

"Who would have thought a gel filtration column would make it so easy for me to get into *Nature*," says Stacy, a post-doc working in New York City.

That may look like just one column, but it's not. You see, that's Superdex™ inside that column. Superdex gel filtration technology is based on the best of the underlying technologies of Sephadex®, Superose® and Sepharose® — each already proven in thousands of published research findings. So, inside a Superdex column, you get the best technologies.

The best qualities of three great technologies in one: Superdex

Superdex delivers extremely rapid separations, incredibly steep selectivity curves, and truly minimal non-specific interactions. So it's ideal for purifying oligonucleotides, peptides, proteins, or other major biomolecules. The ingenious combination of the properties of dextran and agarose makes Superdex ten times faster than comparable gel filtration media. The proof of this statement is defined in the new Pharmacia Biotech "Gel Filtration: Practices and Principles" — a seventh edition handbook (that's right, edition number seven).

We're sure you'll find a Superdex selectivity to satisfy your needs. After all, it's available in three selectivity ranges (prep grade in bulk). And it comes in packs and columns varying from trial-size scouting packs, through lab-to process-scale columns—many of which arrive pre-packed.

Find out more about Superdex. Give us a call: 1 (800) 526-3593 in the USA; +81 3492 6949 in Japan; +46 18 16 50 11 in Europe and the rest of the world. Or visit us on the Internet: <http://www.biotech.pharmacia.se>. We've got the best gel filtration technology solution for you.



**Pharmacia
Biotech**

Uppsala, Sweden. (And the rest of the world)

READER ENQUIRY NO. 715

TB still a global killer

Sir—The World Health Organisation (WHO) recently claimed that the global tuberculosis situation is coming under control thanks to the development of DOTS ('directly observed treatment short-course', which is of 6 months duration)¹. The inference is that tuberculosis could be eliminated by instituting DOTS everywhere, but unfortunately that is far from the reality. There has been a drop of less than 8% in cases of tuberculosis according to WHO, and although affluent countries such as the United States are recording real decreases, many areas of the world such as the Soviet Union are not. Tuberculosis still kills more people each year than any other single infectious disease, including malaria.

There are two main problems. The first is the delivery of medication. Supervised therapy was advocated by physicians at the Hammersmith Hospital in London as long ago as 1958 (ref. 2). Yet, in many parts of the world, adequate infrastructure for DOTS is still not in place and this requires time, commitment and money. DOTS would be simpler to introduce if the effective duration of treatment for tuberculosis was just a matter of weeks, an aim of many researching in the field.

The second but in the longer term more worrying problem is the emergence of drug resistance. *Mycobacterium tuberculosis* resistant to at least one commonly used drug is found in up to 20% of cases in some areas of the world, and multi-drug resistance has been well described. Some of the underlying mechanisms, such as the

role of the catalase-peroxidase system in isoniazid resistance have been elucidated³, but many have not. Second-line drug therapeutic regimens are less effective, more expensive and prolonged.

Now is the time to support research to develop both true short-course therapy (2 weeks, not 6 months) and new immunotherapeutic approaches. If both the difficulties with DOTS and the facts of emerging drug resistance are ignored, the potential scale of the consequent disaster is large, given that *M. tuberculosis* infects approximately one third of the world's population, according to WHO figures⁴. It would be shortsighted to claim that the problem of tuberculosis is receding when all we may be getting from current infection control strategies is time to prepare for a more virulent pathogen. The situation remains similar to that highlighted in *Nature* five years ago⁵ which stressed the need for adequately funded, focused basic science research to combat the worldwide public health problem of tuberculosis.

Jon S. Friedland

Department of Infectious Diseases,
Royal Postgraduate Medical School,
Hammersmith Hospital,
Du Cane Road,
London W12 0NN, UK
e-mail: j.friedland@rpms.ac.uk

1. *Nature* **386**, 534 (1997).

2. Stradling, S. & Poole, G. *Lancet* **ii**, 1066–1067 (1958).

3. Zhang, Y. *et al.* *Nature* **358**, 591–593 (1992).

4. Ravilione, M. C., Snider, D. E. & Kochi, A. *J. Am. Med. Assoc.* **273**, 220–226 (1995).

5. Culliton, B. *Nature* **356**, 473 (1992).

Taking exception

Sir—"It's the exception that proves the rule" is a phrase that is increasingly abused. It actually means that it's the exception that puts the rule to test, that is, studying exceptions leads to refinement and greater understanding of the rule or, alternatively, to the rule being discredited.

Unfortunately, in late twentieth-century usage, the meaning is often that it's the exception that confirms the rule, that is, any rule worth its salt ought to have a few exceptions. That is pseudoscientific, illogical, sloppy thinking, an excuse for not explaining the exception and the exact opposite of the original meaning. What, one asks, would be the consequence of applying this line of thinking to gravitation? Apples flying skyward from the tree would perhaps merely excite mild surprise as scarce exceptions to the law of universal gravitation.

I was therefore appalled to come across this phrase in a News and Views article in

Nature (K. W. Plaxco and M. Groß, **386**, 657–659; 1997). The authors, discussing the finding of an exception to the dogma that the biological active state of a protein is always the folded state, say: "It is not clear how far these findings can be generalized, but certain other proteins show a similar behaviour, suggesting that FlgM is not the exception that proves the rule". But this is precisely what it is, an exception, the first of many, that demands revision of a rule.

When an expression has become so corrupt, it is time for the scientific community to disown it. The expression has already been mistranslated into Norwegian, so I fear that we must surrender it to the vocabulary of scientifically camouflaged superstition, along with 'the law of averages', and find a less ambiguous alternative.

Andrew Jenkins

A/S Telelab,
Strømdalsjordet 4,
3725 Skien, Norway
e-mail: andrew.jenkins@telelab.no

Not a good example

Sir—Harvard University has decided *ex cathedra* that scientific misconduct is not to be attributed to a researcher who has no experimental evidence to support statements made in published papers (*Nature* **386**, 206; 1997). Bertrand Russell once remarked in a similar context that such a decision offers all the advantages that thievery offers over honest toil.

Erwin Klingsberg

Apt 930,
4000 Massachusetts Avenue NW,
Washington, DC 20016, USA

Energy amplifier could have military uses

Sir—Your interesting account¹ of Claude Birraux's preliminary report of the November 1996 hearing at the French parliamentary office of technology assessment on Carlo Rubbia's particle-accelerator-based energy amplifier (the 'Rubbiatron') appears to overlook some essential aspects of the discussion that can be found in Birraux's report.

One of the most animated sessions of the full day-long hearing was that on nuclear proliferation. The reason is that particle accelerators are emerging as a very attractive substitute for nuclear reactors when it comes to building high-intensity neutron sources that can be used to breed large quantities of tritium, plutonium and other special materials of military interest².

Indeed, in the most recent annual report of the French Atomic Energy Commission, the only reference to accelerator-based nuclear systems of the Rubbia type appears in the military section³. As you report, "France currently lacks a [high-current] proton accelerator". Such an accelerator would be very useful for breeding tritium^{3,4} and producing antimatter⁵ for military purposes.

So even if the 'Rubbiatron' fails to produce electricity or to dispose of nuclear waste efficiently, it could still provide a cover and a test-bed for developing high-current accelerators with many important military applications.

André Gsponer

Independent Scientific Research Institute,
Box 30,
1211 Geneva 12,
Switzerland

1. Glover, E. *Nature* **386**, 426 (1997).

2. Gsponer, A. *et al.* *Atomkernenergie-Kerntechnik* **43**, 169–173 (1983).

3. *Rapport annuel 1995*, 33, Commissariat à l'Énergie Atomique, July 1996.

4. Macilwain, C. *Nature* **377**, 567 (1995).

5. Gsponer, A. & Hurni, J.-P. *Atomkernenergie-Kerntechnik* **49**, 198–203 (1987); *Nature* **325**, 754 (1987).

Engineering design lessons from Kobe

Bringing old building stock into line with modern standards of earthquake-resistant design is a daunting and expensive task. But the damage caused by the Kobe earthquake shows that doing nothing will be even more costly.

Adrian M. Chandler

Kobe's city administrators, construction engineers and 1.5 million inhabitants are all determined to turn the aftermath of the disaster that struck the city on 17 January 1995 to the city's gain. Already more than two years into an ambitious ten-year redevelopment programme, Kobe is set to become Japan's first redesigned twenty-first century city, and perhaps therefore its safest in terms of the risk of further earthquake damage. The planning brief highlights "disaster prevention living zones": wide streets creating fire breaks and a duplication of the main routes through the city by both road and rail. It outlines a new Kobe airport, accelerated extensions to an embryonic subway system and a port containing steel-plated earthquake-resistant berths reserved for emergency services.

It is to be hoped that the insights gained by engineers from the Kobe earthquake will be instrumental and effective in reducing future earthquake damage and loss of life in Japan and elsewhere. There is little doubt that an earthquake with similar destructive force could, and probably will, hit Tokyo within the next 20 years. A large earthquake near the capital, last experienced in 1923, is now long overdue. Some earthquake engineers argue that, at present, Tokyo is in practice little better prepared than was Kobe, and that a similar close hit would cause enormous devastation and loss of life. Kobe has served as a timely warning, and its lessons must be quickly and effectively learned.

The urgent need to upgrade the older (more vulnerable), badly designed building stock is evident in all the main earthquake zones in the world. This may take the form of strengthening existing structures, or of introducing an additional form of response control that will reduce the vibrations and damage caused by earthquake ground motions. In the past ten years, research into earthquake response control methods has become very diverse, and it might be more effective if funding were to be more narrowly channelled into the exploration and implementation of tried and tested, cost-effective, useful techniques for retrofitting older buildings, such as base isolation and some varieties of energy-absorbing frame devices, as I will discuss here.

Lessons from Kobe must be learned, and quickly, as there is a real danger of research into new technology for earthquake resistance in countries such as Japan and the United States becoming unduly sophisticated

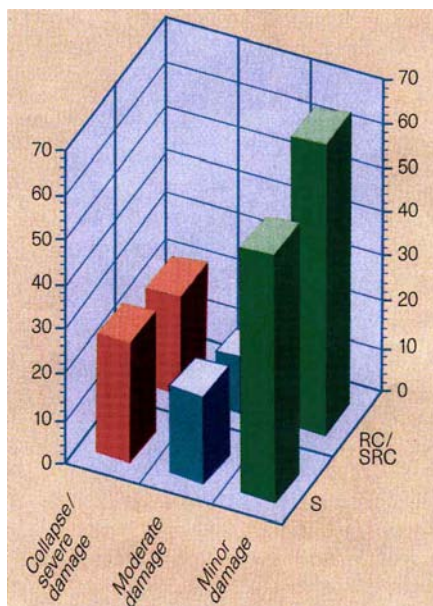


Figure 1 Overall damage statistics in Kobe earthquake, by percentage, for reinforced concrete (RC), steel (S) and composite (SRC) buildings (from ref. 1).

ed and diverse, which may draw attention away from the reality that simple observation of a few basic guidelines can effectively reduce the impact of such events on the built environment.

Kobe earthquake and its aftermath

Early in the morning of 17 January 1995, an earthquake of moment magnitude 6.9 hit south-central Japan, causing considerable casualties and damage. Japan had not experienced an earthquake disaster on such a scale for 46 years, during which time the loss of life owing to building failures in earthquakes was just six people a year. As a result, the extent of damage and loss came as a shock to most people, including structural design engineers and risk-assessment specialists. The total death toll in Kobe exceeded 6,000, with 35,000 people injured. The total direct economic loss resulting from the disaster has been put conservatively at \$130 billion, making Kobe the most costly naturally occurring disaster in human history, and the most devastating ever to hit a developed country. All this happened in about 20 seconds, which was the duration of strong ground shaking for the earthquake. Adding the direct and indirect losses from business disruption and loss of productivity in the city and port is likely to take the final cost to more than \$200 billion.

Virtually all Japan is classified as a medium- or high-risk earthquake zone. In

theory, Kobe is in the same high-risk zone as Tokyo, 420 km to the east. But because no recorded damaging event had hit Kobe for the preceding 1,000 years, earthquake-prediction specialists in Japan rated the chances of such an event as remote, and certainly less than the probability of a repeat of the 1923 great Kanto earthquake, which killed 142,000 people in the capital city.

The 1995 Kobe earthquake was particularly severe because of a combination of three factors. First, it was shallow: the main shock was just 14 km below the surface. Second, the horizontal and vertical ground accelerations were exceptionally strong: these are measured in terms of percentage of g , the acceleration due to gravity, and averaged 50–90 per cent of g horizontally (50–150 per cent greater than conventional design values) and 30 per cent of g vertically over a widespread area, including virtually the whole of Kobe city. Third, the strong motion propagated northeastwards for 50 km from the epicentre (close to Awaji Island, 40 km west-southwest of the city) along a known fault line running through soft alluvial deposits, leading directly to the heart of downtown Kobe. Recorded ground movements highlight the linear band of intense shaking.

More than 56,000 buildings (mainly houses) were totally destroyed by the earthquake, with a further 110,000 severely damaged. The Kobe shock was by far the most devastating earthquake of modern times, measured in terms of damage inflicted on building stock that was, by global standards, constructed to a high standard, designed for appreciable earthquake loadings, and predominantly post-1950 construction. The damage would probably have been significantly heavier had the duration of the earthquake been longer than 20 seconds.

Engineering lessons

What are the implications of the Kobe earthquake as an indicator of the potential for earthquake disaster in other parts of Japan, in particular Tokyo? Despite the particular seismological conditions of the Kobe event (outlined above), it is by no means inconceivable that ground motions of equal severity may strike Tokyo from a larger-magnitude, deeper subduction-zone earthquake, or even from a shallow, moderate-magnitude event similar to Kobe. As in Tokyo, most 'engineered' buildings in Kobe are six to twelve storeys high, with relatively few above fifteen storeys high. These engineered buildings suffered a surprisingly high level of damage¹, with failure or heavy damage

caused to 40–50 per cent of structures made of reinforced concrete, steel or steel/reinforced concrete (composite), as shown in Fig. 1. The use of steel/reinforced concrete as a construction technique for earthquake-resistant buildings, in which steel column sections are encased in a concrete column, is limited mainly to Japan — the standard elsewhere is reinforced concrete, which uses a grid of steel bars to strengthen the concrete columns.

The engineering performance of modern (post-1981) buildings designed to resist a strong earthquake was generally as expected in Kobe, with most surviving the event with minor or moderate levels of damage. The post-1981 Japanese seismic code closely matches the procedures adopted in the United States, Europe and New Zealand, where dense reinforcement in concrete columns is required, including closely spaced steel hoops to encourage ductility (the ability to deform after yield without severe damage or collapse). Hence the most modern Japanese buildings suffered only minor, non-structural damage in the Kobe earthquake, and, with the exception of a single building, did not collapse.

Buildings and transportation structures that collapsed were mostly, but not exclusively, designed and built to less stringent standards than required by modern Japanese codes, reflecting the progressive move towards better earthquake-resistant design technology over the past 30 years or so. For example, recorded damage to reinforced concrete buildings² shows that 57 per cent of those constructed before 1971 suffered collapse or very severe damage; this reduces to 21 per cent for those built between 1971 and 1980, and to near zero for those built to the most modern post-1981 earthquake code standards. The damage to the latter group of structures was estimated overall to have been only one-fifth of that which occurred in the older, pre-1971 buildings.

The damage levels in pre-1971 buildings in the worst-hit parts of Kobe are representative of those to be expected in any future earthquake in a large Japanese or US city, and in other earthquake-prone regions of the world. The proportion of damaged buildings is similar to, or even higher than, that from earthquakes during the past 15 years in Mexico, Armenia, Turkey and other parts of the world that are considered much less well developed than Japan or the United States in terms of earthquake-resistant building technology². This disturbing result explains why there are important questions being asked throughout the world about the adequacy of strengthening and replacement policies for older buildings in earthquake regions³.

The Kobe earthquake, together with that exactly one year earlier in Northridge, California (moment magnitude 6.7), will therefore be remembered as being of special sig-

Figure 2 Severely damaged steel-framed building in Kobe.

nificance largely for three reasons. First, it caused an unexpected level of damage in a part of the world generally thought to be prepared for earthquakes; second, the direct economic damage costs from the Kobe event far exceeded those of any previous earthquakes; and third (but perhaps most important), Kobe has posed questions about the rate at and urgency with which older buildings, clearly much more vulnerable to earthquake damage and collapse, are being replaced or adequately strengthened, in Japan and elsewhere.

Unexpected engineering failures

The age of buildings that failed cannot alone explain the high damage levels to engineered buildings in the Kobe earthquake. There were also some notable collapses and failures in modern steel and concrete high-rise buildings, and these are prompting a radical and far-reaching rethink of design methodologies for earthquake-resistant structures³. There are four important areas being subjected to this critical scrutiny.

First, there was a disturbingly high proportion of failures in steel buildings in Kobe (Figs 1, 2). Damage to welded plate connections in the bracing members of several high-rise buildings was unexpectedly severe, and in several cases buildings could have collapsed if the duration of strong shaking had been a few seconds longer. This type of damage, which could have had catastrophic consequences, echoed the concern over similar types of damage to steel buildings recorded in the 1994 Northridge earthquake⁴. Some mid- and high-rise steel condominium structures incorporated innovative structural systems consisting of steel frames in which the column and girder members were large steel trusses. Damage observed included the brittle fracture of tubular columns and frac-

turing of wide-flange diagonal bracing elements; such damage also occurred in taller, commercial steel structures.

Second, there is cause for concern that this was an earthquake of only moderate magnitude, yet owing to the close proximity of Kobe to the event's epicentre, buildings would have been subjected to forces between 150 and 250 per cent of their designed ultimate resistance. The effect of a much stronger earthquake on one of the large cities of Japan or California could be expected to be yet more devastating⁵. Building codes do not consider explicitly the effect of very close proximity earthquakes (so-called 'direct hits') and may tend, as a consequence, to underestimate the exceedingly high localized ground accelerations close to the fault slip, even in moderate-magnitude earthquakes.

Engineers, particularly from the United States and Europe, are therefore increasingly concerned that existing codes do not fully account for the distinct features of near-field large-magnitude earthquakes. In particular, these are the exceptionally high horizontal ground accelerations, some approaching 100 per cent of *g*; coupling of the horizontal ground accelerations with vertical accelerations that in Kobe were, in some instances, one-and-a-half times the horizontal movement values (these vertical motions are ignored by codes and can lead to failure of columns in buildings and bridges by inducing unaccounted-for tensile forces); and low-frequency 'fling' or 'killer-pulse' ground motions near the earthquake fault, which although of relatively low acceleration can produce excessively large deflections in stiff structures such as elevated bridges and buildings with reinforced concrete structural walls, by acting like a large horizontal static force applied to the structure over a duration

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS
UNAVAILABLE
FOR COPYRIGHT
REASONS

of perhaps one or two seconds, sufficient in many cases to cause the structure to suffer severe damage and in the worst cases to collapse.

Third, the use of steel/reinforced concrete columns in combination with conventional reinforced concrete, as advocated in Japanese building codes, has been brought into question. The former are modern buildings formed of I-section or multiple T-section columns surrounded by mass concrete or reinforced concrete. The main problem appears to lie in the difficulty of connection between the two types of construction, with steel/reinforced concrete generally being used in Japan only up to the sixth-floor level and reinforced concrete construction from the seventh floor upwards. This was probably the main cause of the unusually high number of mid-height failures of modern concrete buildings in Kobe (Fig. 3).

Finally, much of the newer construction in Kobe, particularly of larger buildings, is on very soft, recent alluvial soil and on recently constructed near-shore islands. Most of the serious earthquake damage to larger commercial and industrial buildings and the transportation infrastructure occurred in areas of soft soils and reclaimed land. The worst industrial damage occurred at or near the waterfront, owing to ground failures arising from liquefaction, lateral spreading and settlement. Although the engineering profession has tried hard to develop methods for strengthening filled areas to resist failures during earthquakes, most have been put into practice without the benefit of being adequately tested in strong earthquakes. The performance of these techniques was mixed, but the failures costly — most retaining walls along the port failed, and the consequent ground settlement led to damage of many buildings and port structures.

Innovative structural solutions

Because of the perceived lower risk of earthquakes in Kobe than in Tokyo, only two buildings in the affected region were constructed with base isolators, a technique that uses high-damping rubber bearings to support the columns and thereby reduces the level of forces transmitted to the structure by the ground motions. Both these buildings survived the earthquake undamaged, but given their distance from the epicentre (more than 35 km), neither experienced the full force of the seismic loadings. Nevertheless, designers were encouraged by the observation that the isolators performed exactly as intended, and the buildings were completely in service after the earthquake.

There is great unexploited potential to incorporate this and other innovative construction methods into new buildings for earthquake regions. The additional cost considerably exceeds that of providing conventional earthquake resistance, but may easily

be justified on the basis of improved performance and lower repair costs, particularly for vital structures and facilities such as hospitals and fire stations.

Base isolation is one of an increasing number of methods of modifying structures (including the upgrading or retrofitting of existing buildings, discussed below) to mitigate the disastrous effects of strong earthquakes on buildings and occupants. Many passive-damping (energy-dissipating) devices and active control methods using computer-controlled hydraulic actuators strategically placed in or around the building have been researched in Japan, the United States, New Zealand and elsewhere over the past 20 years. All have their problems, some related to additional costs which may amount to 10–20 per cent of the total construction bill (although compared with potential post-earthquake repair or reinstatement costs for conventional buildings, these are clearly acceptable), whereas others relate to the maintenance and operation of sophisticated methods (particularly active control, which relies on computer-controlled devices), during the lead up to and occurrence of a major earthquake.

Even in Japan, where the industrial research corporations have developed and advanced the technology of these modern methods of earthquake control more, perhaps, than anywhere else, the number of buildings using a form of response control (active or passive) remains small, being considerably less than 1 per cent of the total building stock at risk. At government-funded earthquake research centres in the United States, New Zealand and Europe, there has been a massive expansion of interest in the past 10 years in the development and testing

of both passive and active control methods. What is lacking is a sufficiently widespread will among building owners and developers to implement these techniques and to fund the additional up-front costs or the supplementary costs to upgrade existing vulnerable structures. This may partly be the result of the sheer diversity of methods now available, with no single technique being sufficiently effective or versatile to be considered the clear market leader.

What for the future?

There is a widespread view among engineers that owners and developers must demonstrate greater willingness in the future to pay for proper engineering design and construction to ensure a safe structure. The cost of adequate earthquake resistance is very low — only a few per cent of the building cost — if it is designed into the structure initially, but very high if the building must be upgraded at a later date. In fact, retrofitting an existing building can cost almost as much as constructing a new one. For a large city in Japan or California, the total costs of upgrading all building stock with an effective useful life of upwards of 50 years could be enormous, perhaps \$50–100 billion. Clearly, this cost and the associated construction programme would have to be spread over many years, and could not be funded by conventional government or private sources. Despite this apparent financial deterrent, the lesson from Kobe is that government economists and urban planners must urgently upgrade older, more vulnerable building stock and infrastructure, which has been shown to suffer up to 10 times more earthquake damage than structures built to modern design-code standards in Japan and the United States.

The catastrophic damage and disruption caused in Kobe can be taken as an illustration of the eventual cost of ignoring the problem of older, vulnerable buildings in large cities in earthquake regions. In this context, the expense of upgrading could be regarded as an acceptable price to pay. Engineers have the technology and capability to implement such a programme, but so far the wholehearted political will to go ahead has been lacking. Perhaps Kobe will provide the stimulus for a swing of opinion, but memories are short, and action is required sooner rather than later to prevent a repeat in other earthquake-threatened cities. □

Adrian M. Chandler is in the Department of Civil & Environmental Engineering, University College London, Gower Street, London WC1E 6BT, UK.

1. Elnashai, A. S. *HEI News* No. 15, 6–7 (1996).
2. Earthquake Engineering Field Investigation Team (EEFIT) *The Kobe, Japan Earthquake of 17 January 1995* (Inst. Struct. Eng., London, May 1997).
3. Hayward, D. *New Civil Eng.* 16–20 (13 June 1996).
4. Earthquake Engineering Field Investigation Team (EEFIT) *The Northridge, California Earthquake of 17 January 1994* (Inst. Struct. Eng., London, May 1997).
5. Scawthorn, C. *et al.* *EQE Summary Report* (EQE International, San Francisco, April 1995).

Figure 3 Concrete-frame structure with a mid-height storey collapse.

The value of everything

Stuart L. Pimm

Economists and ecologists have joined forces to estimate the annual value of the services that Earth's ecosystems provide. Most services lie outside the market and are hard to calculate, yet minimum estimates equal or exceed global gross national product.

Economists' self-effacing definition of their craft is that it is one that knows the price of everything and the value of nothing. In that lies a crucial distinction.

Market prices are the easier objects of study, recorded abundantly in units (currencies) with known, if varying, calibrations. Values are more slippery, being likely to vary widely from person to person and from generation to generation. Prices, moreover, reflect incremental (or 'marginal') costs. Diamonds trade for a higher price than does fresh water. Yet the value of all fresh water is infinite: we could not survive without it. For ecologists, prices are woefully incomplete measures of nature's value that poorly prefigure the consequences of humanity's exponential growth. For economists, the challenge is how to assess an ecosystem's non-market values

and predict their future trends. The words 'economics' and 'ecology' share a common root, yet for many years their practitioners had little to do with each other.

A turning point may have been C. W. Clark's 1973 demonstration that the most profitable strategy for harvesting whales is to convert all of them quickly into money in the bank¹. Even the most stingy savings account produces interest income faster than whales reproduce themselves. Using the mutually intelligible medium of mathematical models, Clark's economics explained ecological history. Maximizing profits demanded that the slow-moving, inshore-feeding species were first hunted to extinction, before whalers tackled species whose capture required more advanced technology. The appropriately named right whales did go

first, the 'wrong' whales following at ever shorter intervals. Generally, only species that grow faster than money in the bank should be harvested sustainably.

But if whales are more than just meat, how are we to estimate these other values? Trees in tropical forests grow slowly too: must economics doom them to extinction in the next century, like some whales in the last? On page 253 of this issue², a team of ecologists and economists, sponsored by the newly established National Center for Ecological Analysis and Synthesis in Santa Barbara, California, address such questions. Costanza *et al.*³ take the flippant definition of economics to heart, estimating the annual value of all the world's ecosystems — essentially, the value of everything.

Take whales as an example. Whalers once set out from Lahaina and Nantucket on year-long trips armed with harpoons. Better-paid descendants set out on day-trips, escorting passengers armed with cameras. The cultural value of whales far exceeds that derived from their meat or oil, even if one only counts the price paid for boat tickets. Whales also have ecological roles that maintain species abundances of other marine species, including commercially valuable fisheries. These values, too, could be very large, and easy to estimate in theory, if not in practice.

Estimating other values is more contentious. Many people would consider Baker and Palumbi's observation³ of the presence in Japan's sashimi trade of meat from the long-protected humpback whale to be a violation not only of international laws, but of religious principles, or to be evidence of unacceptable cruelty. Economists incorporate these views by asking how much the public would pay to protect whales. (A real example of such calculations involved asking how much the public felt deprived when the oil from the *Exxon Valdez* gummed up Alaska's scenically spectacular shoreline.)

Herein lies a vigorous debate, for others deem such calculations irrelevant — a feeble, last gasp of economists to fix their inability to assess real values. But would anyone accept child labour just because willingness-to-pay estimates for its abolition were smaller than the money saved by paying children less than adults? Aren't there overarching moral issues in placing monetary values on a sustainable environment for future generations? Costanza *et al.* point out that moral arguments make the discussions more difficult, but go on to say that they have no choice but to try to take them into account. In practice, we do indeed make financial trade-offs involving moral dilemmas, and the authors recommend that we proceed with both moral arguments and economic assessments.

Ecologists must grapple with exceedingly complex ecological interactions. Clear-cut a forest and the price of adjacent homes will

Estimating the cultural value of the oceans

Costanza and colleagues' summary table, Table 2 (page 256), is both fascinating and frustrating: Some numbers are small, others are huge. For none of them is there an indication of how they are obtained. How could there be in a journal with strict space limits? *Nature's* Web site <http://www.nature.com> provides a solution that until recently would have been impossible: it displays the six-page spreadsheet and accompanying 18 pages of footnotes to document the calculations.

For example, how do the authors obtain a value of \$76 per hectare for the cultural value of the world's open oceans? Multiplied

across the planet's oceans this is a huge number — roughly \$2.8 trillion. The tangible economic evidence of valuing the sea is how much more we will pay for coastal real estate than for comparable properties inland. (The calculation for the cultural value of coastal waters themselves is separate, and takes into account such factors as the scientific value of estuaries.)

For California, the difference between coastal and inland real estate is \$10 million per hectare, for Alabama, only \$500,000. Costanza *et al.* estimate the length of the coastlines of wealthy and less developed nations to be 194,435 km and 284,795 km, respectively, and assume coastal

properties extend 0.5 km inland. They further assume that wealthy nations value their coasts 100 times as much as poorer ones, making the latter's contribution relatively tiny.

The total values come to \$5–105 trillion, using the Alabama and California estimates, respectively. Amortized over 20 years, this yields the average value of \$2.8 trillion, or \$76 for each of the oceans' 36.3 billion hectares. All that is before we've added the money spent on tall ships (and yachts), the sextants to sight the stars to steer them by, and all the other paraphernalia needed to mess about in boats. **S. L. P.**

drop. The nearby streams will fill with sediment and lose their fish. Flowing to the sea, the sediment may smother the coral reef offshore, destroying its fish, beautiful corals, and the obscure invertebrate containing clues to the cure for cancer. The forest is a carbon sink. Its loss may accelerate global warming and sea-level rises. The coral reef may provide inshore areas with protection against storms. Costanza and colleagues' approach to these diverse services is to group them into 17 categories (the value of gas regulation, then climate regulation, through to cultural value) and to classify the planet's surface into 16 subdivisions or biomes. Then they proceed to estimate the 272 combinations one at a time. A summary of their valuations appears as Table 2 on page 256 of this issue, a much more detailed version being available as Supplementary Information on *Nature's* Web site (see box on the preceding page for details).

Many combinations have yet to be estimated. Mountains, the Arctic and deserts certainly have cultural values and devoted eco-tourists, but these are absent from the grand total of \$33 trillion per year. Absent, too, are urban areas with city parks. A recent visit to New York's Central Park convinced me that children playing there appreciated its greenery, even if their parents living in the surrounding high-rise housing were too poor to own trees of their own. Filling in these and the other omissions will prove a rich training ground for graduate students in ecological economics.

Some of the numbers will be controversial. If the oceans were not there, re-creating their nutrient cycling would require removing the nutrients from the land's runoff and returning them. The estimate of this service's \$17 trillion value is arrived at by multiplying the cost of removing phosphorus and nitrogen from a litre of waste water by the 40,000 cubic kilometres of water that flow from the land each year (see Web site). It is geological cycles that return these nutrients. In the short term, many would not notice (and perhaps not care) what happens to the elements as they flow into the ocean. Even without this service, the remaining ecosystem services total \$16 trillion annually, a number to compare with the global GNP (gross national product) of \$18 trillion.

Parallel arguments for freshwater ecosystems are more immediate. We use fresh water directly and the function of natural ecosystems in waste treatment is large and frequently overlooked. Here the ecologists' intuition that current prices are a poor predictor of future value is self-evident. The price of an essential resource such as water will rise nonlinearly as it becomes scarce. We already use half of the available global supply of fresh water⁴. In some places, the situation is critical. Israel already uses more water than is available, taking the remainder from

aquifers under the occupied West Bank, and off watersheds in the Golan Heights and southern Lebanon⁵. The region's population is increasing rapidly and its peoples have aspirations for a better life. The nonlinearities in the value of ecosystem services have political as well as economic consequences.

Table 2 on page 256 informs a strategic debate about why we should protect ecosystems. A powerful argument is that we should protect biodiversity as a potential source of new medicines and genes for crop improvement. Costanza *et al.*, quoting as-yet unpublished work by D. Pimentel, show that cultural and recreational uses of nature are even more valuable. Pimentel estimates the value of over-the-counter, plant-based drugs at \$84 billion annually, but eco-tourism at \$500 billion.

The real power of Table 2 lies in its use for local decisions. For example, a proposed project in southwestern Brazil and adjacent countries would straighten and deepen the rivers and so drain many of the surrounding

wetlands — the Pantanal. The value of the increased quantities of soybeans that will be marketed is a simple calculation, but a grossly incomplete evaluation of the project's consequences. The Pantanal is an area that provides unusually high ecosystem services. Even if all the globally averaged benefits do not apply to this particular region, the global accounting provides a checklist for those who make difficult local decisions. Increasingly, such decisions will be informed by those who realize that there is more to a whale than its meat, and that wetlands, like all other ecosystems, provide services we cannot afford to replicate. □

Stuart L. Pimm is in the Department of Ecology and Evolutionary Biology, The University of Tennessee, Knoxville, Tennessee 37996, USA.

1. Clark, C. W. *Science* **255**, 890–897 (1973).
2. Costanza, R. *et al.* *Nature* **387**, 253–260 (1997).
3. Baker, C. S. & Palumbi, S. *Science* **265**, 1538–1539 (1994).
4. Postel, S. L., Daily, G. C. & Ehrlich, P. R. *Science* **271**, 785–788 (1996).
5. Postel, S. L. *Last Oasis: Facing Water Scarcity* (Norton, New York, 1992).

Geomorphology

When streams collide

Chris Paola

Dramatic effects can occur when two solid objects collide. The effects produced when two rivers collide are no less dramatic, but they are less familiar because they mostly happen under water.

The veil is lifted for one of the largest rivers in the world in the paper by Best and Ashworth on page 275 of this issue¹, which reports a sequence of bathymetric surveys in and around the confluence of the Jamuna and Ganges rivers in Bangladesh over a period of 28 months. They have discovered one of the mightiest natural scour holes ever described: about 30 m deep, some five times deeper than the channels feeding into it. Despite its great size, the scour is nimble: it migrated downstream at a mean rate of nearly 2 km per year over the measuring period.

There are several perspectives from which the size and dynamism of this impressive hole in the river bed are important. But first, why is the scour there, and how does it form^{2–5}? Think of the confluence of two nearly parallel streams with different speeds. As the flows merge, the difference in speed gives rise to velocity shear and formation of the familiar line of vertical-axis vortices that one commonly sees in streams. In contrast, symmetrical confluences with more abrupt angles also produce vortices, but their axes are submerged and roughly flow-parallel (Fig. 1). So they are not obviously visible from the surface. The opposed components of momentum of the two colliding streams push the free surface of the water up in the

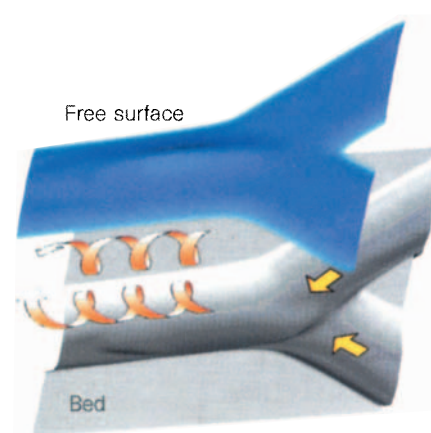


Figure 1 Exploded, schematic view of the bed, flow pattern and free surface at a stream confluence. Collision of the two incoming streams raises the free surface and forces a downwelling flow along the boundary of the paired helical vortices, resulting in a deep scour on the bed and, as discussed here, a scour hole such as that identified by Best and Ashworth¹.

neighbourhood of the junction and force a strong downwelling of water beneath the surface.

As this downwelling water is driven sideways and back up again by the bed, all the while being carried downstream by the mean flow, it wraps back on itself to form a pair of powerful, counter-rotating vortices like those behind a twin-screw power boat. The downwelling region along the boundary of

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Figure 2 Confluence between the Meghna and Padma rivers, also in Bangladesh. The same underwater processes acting at the Jamuna–Ganges confluence apply, but here the high suspended-sediment load of the Padma (milky white) makes the surface mixing highly visible.

these two vortices pulls fast-moving surface water towards the bed. This 'momentum pump' produces very high stresses on the bed, and is primarily responsible for the formation of the deep scour pit. (This is why, even though they are often relatively narrow, confluences are dangerous places to cross streams.) The more abrupt the angle between the colliding streams, the more powerful the streamwise vortex pair. The confluence studied by Best and Ashworth, with its junction angle near 90°, is nicely configured to produce a deep scour.

Confluences are a basic building block of river networks on all scales. They are the main topological element of the dendritic (tributary) networks that cover much of the Earth's landscape; they also form wherever two branches of a braided (multichannel) river rejoin. What effect do they have on river deposits? Stratigraphers have long wrestled with the question of how to separate the 'signal' from the 'noise' in the Earth's sedimentary record: here the signal is the sedimentary signature of some externally imposed change, such as falling sea level; the noise is the signature of the transport system's own internal dynamics (termed 'autocyclicity' in pre-chaos time, when the processes were all assumed to be cyclic). Ideally, one wants a set of minimal criteria that can be used to decide when a signal is clearly not autocyclic, and so is externally forced, and to do that one has to know the limits of what the system is capable of doing on its own.

The scour pattern at the Jamuna–Ganges junction is certainly autocyclic, yet it is so

deep and so mobile that the surface it mills out might easily be mistaken for scour produced by fluvial erosion during sea-level fall (when sea level falls, rivers tend to cut valleys that allow them to maintain a smooth transition between the river surface and the sea; this erosion produces the 'sequence-stratigraphic boundaries' alluded to in the title of the paper). Best and Ashworth propose a more stringent set of criteria for ruling out autocyclicity; one lesson is that it may be better to look for the signatures of modest sea-level changes in small river systems whose internal fluctuations are commensurately modest.

Large, active scours are important in other areas as well. One hates to think what a 30-m deep hole that can move 2 km in one year implies for bridge design. Likewise, tunnels, pipelines and other structures passing under river systems must leave room overhead for the deep bites taken out of the river bed by colliding streams. Confluences also play a disproportionate role in moving sediment through braided river systems, thus helping to determine their total width⁹. Finally, the bases of the fluvial 'ribbon' sandstones that form sinuous, permeable conduits through floodplain deposits are probably shaped in part by these natural milling machines.

The work of Best and Ashworth has shown us how large and mobile confluence scours can be, but we still have much to learn about their kinematics and dynamics. One

way forward is suggested by the observation that the ratio of scour depth to the average depth in the inflow channels in the Jamuna–Ganges system (about five) is consistent with a scaling relation that was developed from experiments in a laboratory model stream with channels of the order of only a centimetre deep¹⁰. This striking geometric similarity over a three-orders-of-magnitude range in depth scale provides a beautiful illustration of scale independence in a fundamental process of river dynamics. It also suggests that there is still much to be gained by complementary experimental and field studies of stream braiding. □

Chris Paola is in the Department of Geology and Geophysics, and the St Anthony Falls Laboratory, University of Minnesota, Minneapolis, Minnesota 55455, USA (e-mail: cpaola@maroon.tc.umn.edu).

1. Best, J. L. & Ashworth, P. J. *Nature* **387**, 275–277 (1997).
2. Ashmore, P. E., Ferguson, R. I., Prestegard, K. L., Ashworth, P. J. & Paola, C. *Earth-surf. Proc. Landfm.* **17**, 299–311 (1992).
3. Best, J. L. in *Recent Developments in Fluvial Sedimentology* (eds Ethridge, F. G., Flores, R. M. & Harvey, M. D.) 27–35 (Soc. Econ. Paleontol. Mineral. 1987).
4. Biron, P., Roy, A. G., Best, J. L. & Boyer, C. J. *Geomorphology* **8**, 115–129 (1993).
5. Mosely, M. P. *J. Geol.* **84**, 535–562 (1976).
6. Rhoads, B. L. & Kenworthy, S. T. *Geomorphology* **11**, 273–293 (1995).
7. Roy, A. G. & Bergeron, N. *Geomorphology* **3**, 99–112 (1990).
8. Bridge, J. S. & Gabel, S. L. *Sedimentology* **39**, 125–142 (1992).
9. Paola, C. in *Coherent Flow Structures in Open Channels* (eds Ashworth, P. J., Bennett, S. J., Best, J. L. & McLelland, S. J.) 705–723 (Wiley, Chichester, 1996).
10. Ashmore, P. & Parker, G. *Wat. Resour. Res.* **19**, 392–402 (1983).

Muscle structure

Molecular bungees

Thomas C. S. Keller III

Molecules of the extraordinarily long and elastic muscle protein titin¹ (or connectin²) work like molecular bungee cords, to passively resist the stretching of relaxed striated (skeletal or cardiac) muscle. The structural basis for this remarkable mechanical property has now been worked out using exciting new methods for grabbing and stretching individual titin molecules. On page 308 of this issue³, Tskhovrebova *et al.* describe the results of a study using optical tweezers, and two papers in *Science* by Kellermayer *et al.*⁴ and Rief *et al.*⁵ come to similar conclusions using optical tweezers and atomic-force microscopy, respectively.

In each half of the thousands of contractile units — sarcomeres — that make up muscle, titin molecules span the >1- μ m distance between the M- and Z-lines (Fig. 1, overleaf). Titin also associates with half of a bipolar myosin (thick) filament, and these amazingly elastic titin connections provide relatively little resistance to the extension that is needed to reset the sarcomere after active contraction. But they increasingly

stiffen as the sarcomere elongates, to resist sarcomere distension beyond the point at which the myosin filaments and actin (thin) filaments can no longer interact to produce contractile force.

If sarcomeres are stretched beyond what is known as the 'yield point' — a degree of stretch that is rarely encountered *in vivo* — titin dissociates from the myosin filaments, and it can extend up to four times its normal length without losing its Z- and M-line anchorages⁶. This fail-safe mechanism maintains the structural integrity of the muscle as a whole, although that sarcomere may be unable to contract, at least temporarily. During normal contraction, the elasticity of titin also helps to balance the contractile forces that are produced by myosin, so these filaments remain centred in the sarcomere⁷.

Clues to understanding titin's mechanical properties came from cloning feats that yielded the complete sequence of titin from human striated muscle⁸, and an invertebrate mini-titin⁹. Much of the ~30,000-amino-acid vertebrate titin molecule folds into tandem arrays and super-repeats of the

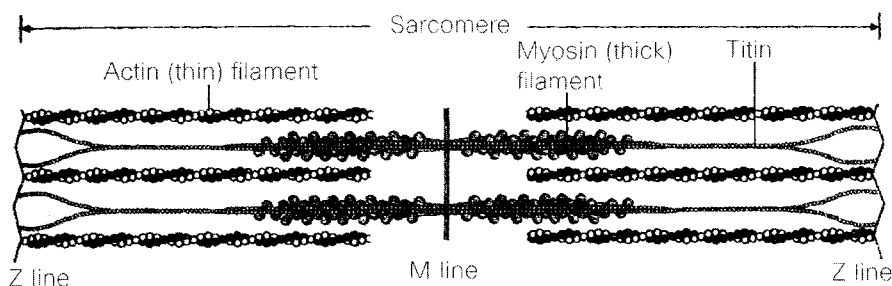


Figure 1 Vertebrate titin molecules span half of a muscle contractile unit (sarcomere) from the M-line to the Z-line; these lines are poorly defined structures to which the myosin and actin filaments are respectively anchored. Only two of the three to six titin molecules that might be associated with each myosin filament in half a sarcomere are shown. Mechanochemical interactions between myosin heads and actin filaments 'walk' the myosin over the actin towards the respective Z-line in each half-sarcomere. New studies by Tskhovrebova *et al.*³, Kellermayer *et al.*⁴ and Rief *et al.*⁵ indicate that muscles passively resist overextension mainly through elastic unfolding of the titin PEVK domain, and unfolding of immunoglobulin domains under extreme stretch conditions. (Modified from ref. 13.)

immunoglobulin-C2 and fibronectin-III structural domains, each of which contains ~100 amino acids. Both of these domains are sandwich-like structures (Fig. 2) comprising seven β -strands, and they are stabilized by hydrogen bonds between the strands that form each sheet, and by side-chain interactions between the two faces of the sandwich.

The bungee-like region of titin between the end of the myosin filament and the Z-line is composed almost entirely of immunoglobulin domains (70–90, depending on the isoform) flanking a unique region of 1,000–2,000 amino acids, named the PEVK segment⁸ because it is rich in proline (P), glutamic acid (E), valine (V) and lysine (K). Using antibody binding to monitor specific regions of the titin molecule, the PEVK segment has clearly been shown to extend reversibly in sarcomeres that are stretched to varying degrees¹⁰. At greater issue has been the part in stretch resistance played by unfolding of the immunoglobulin and fibronectin-III domains¹¹. The new studies

on the visco-elastic properties of isolated titin molecules^{3–5} confirm that the immunoglobulin domains unfold under extreme stretch conditions, and they also reveal considerable differences in elasticity between the PEVK segment and immunoglobulin domains.

Tskhovrebova *et al.*³ and Kellermayer *et al.*⁴ tethered titin molecules between a moveable substrate and a bead captured in an optical trap, and then stretched them. Both studies indicate that titin extends in two distinct phases. As the titin molecule is stretched, the first bungee-like phase of extension contributes little resistance as the fully folded form of titin straightens. But additional stretch generates exponentially increasing levels of resistance, as the PEVK segment unfolds to an extended polypeptide chain about 0.4–0.8- μ m long — greater than ten times its folded length. This PEVK segment readily and rapidly refolds as the titin is released.

When the first component is fully ex-

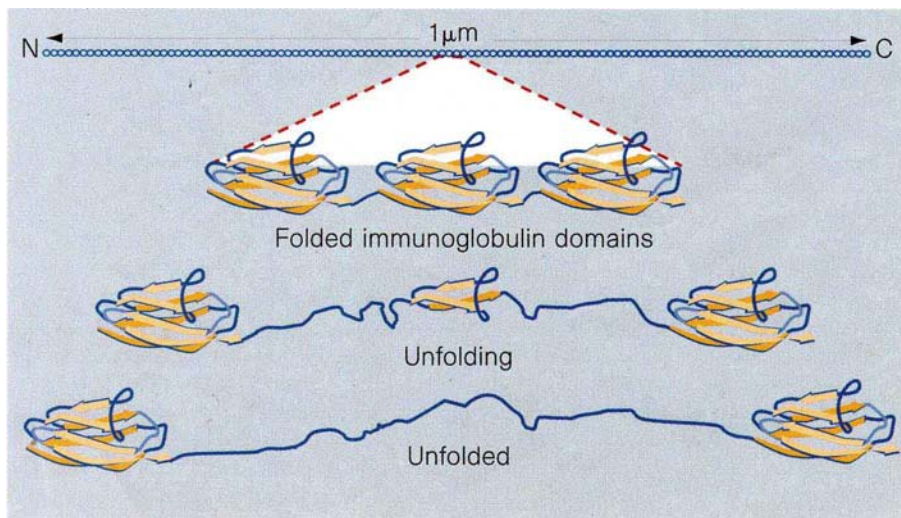


Figure 2 Structure of an immunoglobulin domain, folded and unfolded. The titin molecule stretches due to reversible disruption of the non-covalent bonds that stabilize the structure of the folded domain. (Modified from ref. 13.)

tended, more extreme stretching brings the second, less elastic component into play: the individual immunoglobulin (or fibronectin-III) domains are rapidly unfolded in an all-or-none fashion, through disruption of the stabilizing bonds (Fig. 2). Refolding of these domains is less elastic, because it is relatively slow, and only occurs when the domains are subjected to little, or no, stretch force (do not jump with a bungee that does this!).

Using atomic-force microscopy, Rief *et al.*⁵ characterized the stretching behaviour of individual immunoglobulin domains in short segments of four or eight bacterially expressed titin domains. They found that each domain unfolds in the same way as in intact titin (that is, all-or-none) but, like the extension of a shock absorber, the force necessary for domain unfolding depends on the extension speed — the faster the extension, the greater the resistance to unfolding. This study also showed that there is a clear and persistent hierarchy of unfolding, in which some domains consistently unfold at a lower stretch force than others. So minor differences in amino-acid sequence — especially in the identity of the specific amino acids that hold the faces of a given domain together — may determine the specific role of that domain in titin elasticity.

This possibility is particularly intriguing when coupled with an observation made by Kellermayer *et al.*⁴. They found that repetitive stretching cycles lead to progressive levels of denaturation in which individual domains fail to refold on release. They attribute this to proline isomerization, which can occur in the unfolded stage. If such domains remain unfolded in muscle, they could effectively increase the range of stretch over which resistance is low, and such physical conditioning at the molecular level may allow muscles to adapt to changing conditions.

The three new studies^{3–5} give an insight into the mechanical properties of a structural molecule and foretell that there is more to learn, not only about the structure and function of titin in striated muscle, but in cytoskeletal structures of non-muscle cells, where a cellular version of titin has been found¹². They also provide a new jumping-off point for understanding unfolding and refolding in general — immunoglobulin and fibronectin-III domains are found in a variety of proteins. In extracellular proteins, many of the domains are further stabilized by disulphide bonds (which prevent unfolding) between conserved cysteine residues in different faces of the β -sheet sandwich. Similar domains in intracellular proteins such as titin lack these conserved cysteines, so they may be structurally less stable. Because these domains often mediate the interactions between proteins — as is the case for titin and myosin — their mechanical properties raise

the possibility that stretch-induced unfolding can also regulate protein-protein interactions. □

Thomas C. S. Keller III is in the Department of Biological Science, Florida State University, Tallahassee, Florida 32306-3050, USA.

1. Wang, K., McClure, J. & Tu, A. *Proc. Natl Acad. Sci. USA* **76**, 3698–3702 (1979).
2. Maruyama, K. *et al. J. Biochem. (Tokyo)* **82**, 317–337 (1977).
3. Tskhovrebova, L., Trinick, J., Sleep, J. A. & Simmons, R. M. *Nature* **387**, 308–312 (1997).
4. Kellermayer, M. S. V., Smith, S. B., Granzier, H. L. & Bustamante, C. *Science* **276**, 1112–1116 (1997).

5. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M. & Gaub, H. E. *Science* **276**, 1109–1112 (1997).
6. Wang, K., McCarter, R., Wright, J., Beverly, J. & Ramirez-Mitchell, R. *Biophys. J.* **64**, 1161–1177 (1993).
7. Horowitz, R. & Podolsky, R. J. *J. Cell Biol.* **105**, 2217–2223 (1987).
8. Labeit, S. & Kolmerer, B. *Science* **270**, 293–296 (1995).
9. Benian, G. M., L'Hernault, S. W. & Morris, M. E. *Genetics* **134**, 1097–1104 (1993).
10. Linke, W. A. *et al. J. Mol. Biol.* **261**, 62–71 (1996).
11. Erickson, H. P. *Proc. Natl Acad. Sci. USA* **91**, 10114–10118 (1994).
12. Eilertsen, K. J., Kazmierski, S. T. & Keller, T. C. S. III *J. Cell Biol.* **126**, 1201–1210 (1994).
13. Keller, T. C. S. III *Curr. Opin. Cell Biol.* **7**, 32–38 (1995).

Condensed-matter physics

Organic superconductivity grows in high fields

Denis Jérôme

Large magnetic fields destroy superconductivity. But Lee and colleagues¹, in the 5 May issue of *Physical Review Letters*, have found that superconductivity in a certain organic salt is very resilient to the application of large magnetic fields parallel to one crystallographic direction, and might eventually be stabilized by a high enough field. Superconductivity in their material can already survive fields of around six tesla, whereas it would be destroyed at less than two tesla in any other superconductor with a similar critical temperature.

In the early 1980s, some organic molecular materials were found to be superconducting (that is, have zero resistance and expel magnetic fields). To a large extent the search for superconductivity in organic materials had been boosted by a suggestion²

that the attraction between electrons in Cooper pairs — a necessary ingredient in the conventional theory of superconductivity, the Bardeen, Cooper and Schrieffer model — could be of an unfamiliar nature: instead of an attraction mediated by the interaction of individual electrons with the surrounding lattice, an attractive force could result from the interaction between the paired electrons and the electrons in highly polarizable molecules.

According to Lee *et al.*¹, the high-field superconductivity stems from the particular design of their materials, which are made of flat, conjugated fulvalene molecules (in full, the Bechgaard salt (TMTSF)₂PF₆, or tetramethyltetraseleno-fulvalene hexafluorophosphate), packed in columns along the *a*-direction (Fig. 1). This direction has a

high, metallic conductivity³; the transverse directions have much weaker intermolecular coupling, and thus a much lower conductivity.

In all superconductors, superconductivity begins below a critical temperature when pair condensation lowers the overall free energy of the electron gas, making the paired superconducting phase more stable than the ordinary metallic phase. But the paired electrons acquire extra kinetic energy in a magnetic field, which can destroy the superconducting ground state.

In type-II superconductors, superconductivity survives the application of a magnetic field up to the orbital critical field H_{c2} , where the increase of kinetic energy balances the pair condensation energy. But even if the field H_{c2} is very large in a particular material, there is usually another limit to the stability of superconductivity: pairing almost always takes place between electrons with antiparallel spins, but electron spins tend to become aligned parallel in a magnetic field. Equivalently, pairs are destabilized by the increase of paramagnetic energy due to a large enough magnetic field; this is the paramagnetic limiting field H_p ($H_p = 1.84 T_c$, where T_c is the critical temperature in kelvin).

Lee *et al.*¹ have studied superconductivity in (TMTSF)₂PF₆ down to very low temperatures, keeping it at high pressure (otherwise the stable low-pressure phase is insulating and magnetic). The orbital critical field behaves oddly: parallel to the *a*- or *b*-axis, H_{c2} increases more and more rapidly with decreasing temperature, and as the temperature approaches zero it overcomes the Pauli limit H_p by at least a factor of two. According to the authors, this had not been revealed before because a near-perfect alignment (within 0.1°) is needed between the magnetic field and the *a*- or *b*-crystal axis.

Magnetic fields are known to play a decisive role in the low-temperature stability of quasi-one-dimensional conductors. When the field is applied along the direction of weakest conduction (the *c*-axis), superconductivity is destroyed by the increased pair kinetic energy, and a series of antiferromagnetic phases are stabilized^{4,5}. In this geometry, the electrons oscillate in a plane perpendicular to the magnetic field orientation, with an amplitude inversely proportional to the field strength. Thus, the magnetic field may reduce the effective intercolumnar coupling, and so inhibit transverse hopping⁶. A magnetic field along the *b*-axis of a quasi-one-dimensional anisotropic conductor should inhibit electron motion in the *c*-direction. Lebed predicted⁷ that because of this confinement of electrons to the *a*-*b* plane, superconductivity should return at very high magnetic fields in the case of parallel spin pairing.

In fact, the phase diagram is more complicated⁸: increasing the field leads to a series of temporary superconducting phases, each

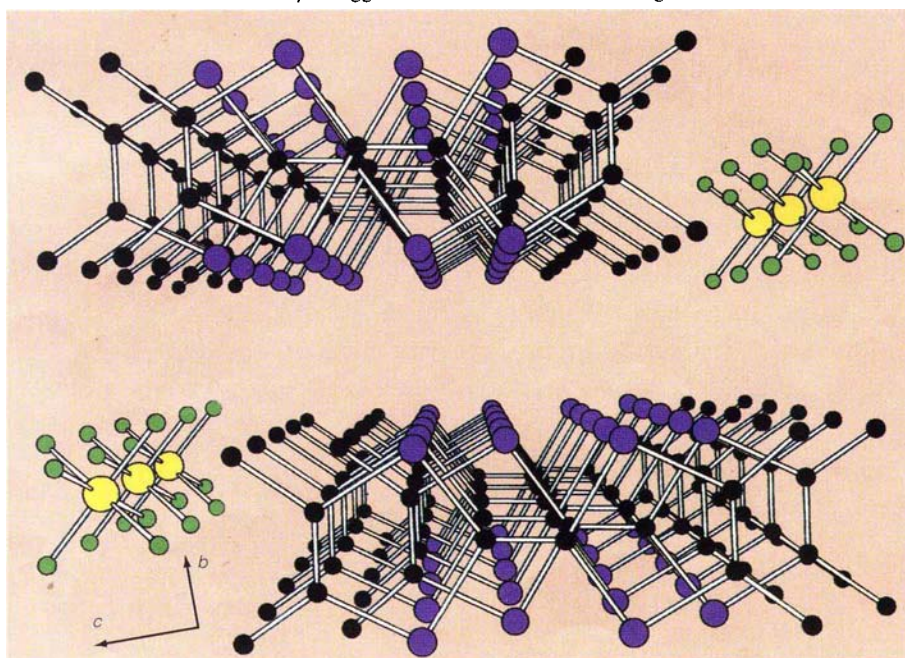


Figure 1 Fulvalene molecules and hexafluorophosphate anions stack along the *a*-axis in the TMTSF₂PF₆ organic superconductor, which has an exceptionally high resilience against magnetic fields parallel to the *a*- or *b*-axis.

for a particular correspondence between the lattice spacing in the *c*-direction and the period of the pair wavefunction. In the extreme high-field limit, the superconducting phase should return with a critical temperature increasing, instead of decreasing, with the magnetic field. This spectacular high-field 're-entrance' mostly involves electrons pairing with parallel instead of antiparallel spins.

Lee *et al.*¹ have stressed that the survival of superconductivity at magnetic fields exceeding by a large amount the paramagnetic limit suggests parallel spin pairing. In fact, by measuring the temperature at which the resistivity reaches zero under a magnetic field, they are detecting the melting of a lattice of magnetic vortices, rather than the temperature dependence of H_{c2} . But their conclusion that $H_{c2} \gg H_p$ is still valid.

A lot of work must be done before we can unambiguously claim that these materials show the re-entrance of superconductivity

predicted for the high-field regime. In particular, higher fields (above 20 tesla) and lower-temperature experiments with accurate field alignment should be feasible in the near future.

We do not foresee very high critical temperatures in quasi-one-dimensional superconductors, so it isn't clear what applications they might have. But they are important materials for testing our models of superconductivity and, more generally, the theory of electrons in low-dimensional spaces. □

Denis Jérôme is in the Laboratoire de Physique des Solides, Université Paris-Sud, 91405 Orsay, France.

1. Lee, I. J., Naughton, M. J., Danner, G. M. & Chaikin, P. M. *Phys. Rev. Lett.* **78**, 3555–3558 (1997).
2. Little, W. A. *Phys. Rev. A* **134**, 1416–1424 (1964).
3. Bechgaard, K. *et al. J. Am. Chem. Soc.* **103**, 2440–2442 (1981).
4. Cooper, J. T. *et al. Phys. Rev. Lett.* **63**, 1984–1987 (1989).
5. Hannahs, S. T. *et al. Phys. Rev. Lett.* **63**, 1988–1991 (1989).
6. Gork'ov, L. P. & Lebed, A. G. *J. Phys. (Paris)* **45**, L433–L440 (1984).
7. Lebed, A. G. *JETP Lett.* **44**, 114–117 (1986).
8. Dupuis, N., Montambaux, G. & Sa de Melo, C. A. R. *Phys. Rev. Lett.* **70**, 2613–2616 (1993).

Oceanography

The limits to growth

Louis A. Codispoti

That the availability of nutrients and light controls carbon fixation by oceanic plants is a proposition that attracts little debate. But there has been hot discussion over just what nutrients limit this process and the applicable timescales. The traditional argument has centred around the relative importance of phosphorus and nitrogen¹. More recently, iron and other trace metals have received considerable attention, and experiments² have shown that carbon fixation in some parts of the ocean can be limited by iron.

Most investigators agree that fixed nitrogen can be limiting over shorter time intervals, but those who favour phosphorus as the ultimate controlling nutrient over periods of more than 10,000 years may feel their hackles rise when they read Paul Falkowski's latest paper on page 272 of this issue³. Falkowski asserts that it is nitrogen, not phosphorus, that limits primary productivity on geological timescales. He bases his arguments, in part, on a discussion of the evolutionary history of the enzyme systems involved in nitrogen fixation and denitrification which seem to be the main source and sink, respectively, for oceanic fixed nitrogen.

Denitrification seems to have evolved independently several times, leading to diversity in the enzymes and microbes responsible. This is not so for nitrogen fixation and, to quote Falkowski, "...the sequence of the genes encoding the catalytic subunits for nitrogenase [the nitrogen-fixing enzyme] is highly conserved in cyanobacteria and other eubacteria, strongly suggesting

an ancient, common ancestral origin".

A key point is that nitrogenase, which seems to have originated before the atmosphere was well oxygenated, has requirements for iron and for anoxia that are not well-matched to present-day oceanic conditions.

Jupiter's moons

Virgin Callisto

The nymph Callisto wanted to remain a virgin, like her hunting companion

Artemis. Inevitably, she was seduced by Zeus (or Jupiter), and then got turned into a bear for her pains by the jealous Hera. But sometimes the real world is kinder than myth: the Galileo spacecraft has discovered that Jupiter's outermost large moon (right) is a virgin aggregate of rock and ice.

The difference between Callisto and its siblings is striking. Unlike Io, Europa and Ganymede, Callisto shows no evidence of an internal magnetic field (D. A. Gurnett *et al.* and K. K. Khurana *et al. Nature* **387**, 261 and 262; 1997). In the other moons, the field probably comes from a molten iron core, which

formed when the moon got hot enough for ice, rock and metal to melt and separate. And indeed, Callisto's gravitational field shows that, unlike the other moons, it has a fairly uniform density, with maybe just a thin surface ice layer (J. D. Anderson *et al. Nature* **387**, 264; 1997).

Why is Callisto so different from Ganymede, which is almost the same size? Early heating of the two moons, from accretion and internal radioactivity, would have been about the same. But Ganymede may once have passed through an orbital resonance, where it was melted by tidal heating from Jupiter; Callisto, unlike her namesake, appears to have avoided such warm embraces.

Stephen Battersby



hemipelagic sediments. They are also present in suboxic portions of the oceanic water column, which occur at depths of a few hundred metres, and are persistent features in a few areas and occur sporadically in others. In addition, the denitrification rate is sensitive to small redistributions of dissolved oxygen and organic-matter fluxes, and to glacial/interglacial changes in continental-shelf area^{1,7}. Because the sites of increased nitrogen fixation and denitrification are different, and because these processes respond to different forcing, imbalances should be expected at least on timescales shorter than the oceanic circulation time of about 1,000 years¹.

An example of the difficulty for nitrogen fixation in today's ocean is provided by the Arabian Sea. Denitrification in the suboxic portion of this region's oxygen minimum zone¹ helps to provide waters that have low ratios of inorganic fixed nitrogen (nitrate, nitrite and ammonium) to phosphate to the sunlit upper layers. In addition, high rates of nitrogen fixation are favoured by a high iron supply⁸. Nevertheless, a search of the extensive, high-quality data on inorganic nitrogen and phosphate collected by myself and colleagues during a complete annual cycle in the Arabian Sea has, to date, found no evidence for sufficient nitrogen fixation to redress the effects of denitrification. Relative to the ~16:1 fixed-N:P ratio (by N and P atoms) of uptake during carbon fixation³, the prevailing ratios are quite low and, in surface waters with low nutrient levels, phosphate was more abundant (relative to the 16:1 ratio). The point is that it takes more than a simple increase in the iron supply to redress deficiencies in fixed nitrogen, and details of both the supply and demand sides must be carefully considered.

Perhaps a subliminal reason for the resistance to the notion of nitrogen rather than phosphorus limitation is that acceptance of the potential importance of the coupling between nitrogen fixation and denitrification, and its relationship to fluxes of iron, organic matter and oxygen, plunges us into new levels of complexity in attempting to predict changes in the ocean's ability to fix carbon. The overall turnover rate between nitrogen fixation and denitrification in the sea is on the order of 100 Tg N yr⁻¹ (Tg = 10¹² g), with my best guess for the denitrification rate being 200 Tg N yr⁻¹ or more. Falkowski points out that an imbalance between nitrogen fixation and denitrification of 3 Tg N yr⁻¹ can largely explain changes in the atmospheric carbon dioxide record over the last interglacial/glacial-maximum cycle, so the sensitivity of the system is frightening. Moreover, we must add the vagaries of the aerial iron supply that seems to be the main source of iron for the open ocean. This supply is very unevenly distributed and dominated, in some regions at least, by weather conditions prevailing over periods of only a

few days⁸. If we add oceanic nitrous oxide and other complications to this picture, the situation becomes even more daunting⁹.

A cautionary note that emerges is that simple 'solutions' for problems in such a complex system are likely to have unintended consequences. Take the idea of iron fertilization, which might superficially be looked upon as 'good' because it should increase carbon fixation² and nitrogen fixation. There is debate about the efficiency of iron fertilization, but, if it has a significant effect, the increased downwards flux of organic material might increase the nitrous oxide flux from the ocean⁹, and increase denitrification (remember that the compensating process, nitrogen fixation, requires about a hundred times more iron than does carbon fixation). Other deleterious outcomes are also possible⁹. Along with many colleagues¹⁰, I fear how governments might simplistically apply our increasing knowledge of the 'leverage' of iron and other trace chemicals to oceanic biology. Applying simple solutions to such a poorly understood, poorly coupled

and highly sensitive system is likely to be a recipe for disaster.

Although the complexity of the system is indeed daunting, can we continue to avoid paying more attention to the processes that influence the nitrogen fixation/denitrification couple? □

Louis A. Codispoti is in the Center for Coastal Physical Oceanography (CCPO), Old Dominion University, Norfolk, Virginia 23529, USA.

1. Codispoti, L. A. in *Productivity of the Oceans: Present and Past* (Berger, W. H., Smetacek, V. S. & Wefer, G., eds) 377–394 (Wiley, Chichester, 1989).
2. Coale, K. H. *et al.* *Nature* **383**, 495–501 (1996).
3. Falkowski, P. G. *Nature* **387**, 272–275 (1997).
4. Paerl, H. W., Pinckney, I. L. & Kucera, S. A. *Limnol. Oceanogr.* **40**, 634–638 (1995).
5. Howarth, R. W., Swaney, D., Marino, R., Butler, T. & Chu, C. R. *Limnol. Oceanogr.* **40**, 639–642 (1995).
6. Piper, D. Z. & Codispoti, L. A. *Science* **188**, 15–18 (1975).
7. Christensen, J. P., Murray, J. W., Devol, A. H. & Codispoti, L. A. *Global Biogeochem. Cycles* **1**, 97–116 (1987).
8. Prospero, J. M. in *Particle Flux in the Ocean* (eds Ittekkot, V., Schafer, P., Honjo, S. & Depetris, P. J.) 18–52 (Wiley, Chichester, 1996).
9. Fuhrman, J. A. & Capone, D. G. *Limnol. Oceanogr.* **36**, 1951–1959 (1991).
10. Chisholm, S. W. & Morel, F. M. M. (eds) *Limnol. Oceanogr.* **36**, 1–1970 (1991).

Synthetic chemistry

Cancer drugs better than taxol?

Cameron J. Cowden and Ian Paterson

The clinical and commercial success of taxol* (1, Fig. 1) in cancer chemotherapy has stimulated a worldwide search for compounds with a similar mode of action but superior properties — particularly activity against drug-resistant tumours. The epothilones are a new class of natural cytotoxic compounds produced by a cultured strain of the myxobacterium *Sorangium cellulosum*, which are attracting increasing attention and excitement from chemists, biologists and clinicians^{1,2}. They work in the same way as taxol to stop cancer cells from proliferating³, and based on initial *in vitro* studies, they may be better drugs, in that they are active against some taxol-resistant cancer-cell lines. In the past six months, no fewer than five total syntheses of epothilone A (2) have been reported (from the groups of Danishefsky, Nicolaou and Schinzer)⁴. To add to these exciting results, Nicolaou *et al.* (page 268 of this issue⁵) report the synthesis of both epothilones A (2) and B (3), as well as an epothilone analogue 4 with greater tubulin-assembly activity than either of the naturally occurring epothilones or taxol.

The microtubule cytoskeleton plays an important part in cell mitosis, which relies on polymerization and depolymerization of the protein tubulin. Many established anti-cancer agents such as the vinca alkaloids and colchicine work by inhibiting microtubule assembly; taxol is unusual in that it pro-

motes the formation of stable bundles of microtubules, killing the cell by inhibiting microtubule disassembly. Despite having little structural similarity, the epothilones as well as discodermolide^{6,7} (5 in the figure), work in the same way as taxol. In fact, the epothilones and discodermolide competitively inhibit the binding of taxol to tubulin polymers, probably indicating overlap between their respective binding sites.

Of these compounds, the epothilones are structurally the simplest, and hence amenable to the synthesis of analogues. With a range of analogues, relationships between structure and activity can be found. The structures of the epothilones (2 and 3) include a 16-membered macrocyclic lactone, and differ only in the presence of a hydrogen or methyl group at C12. They both have seven chiral centres with a range of functionality, including a thiazole-containing side chain and an epoxide. Although the epothilones can already be made by fermentation, chemical synthesis will allow access to a wide range of structural analogues.

The new total synthesis⁵ of epothilone A is notable in that the initial steps are undertaken on a substrate attached to a resin (Fig. 2). Such solid-phase chemistry is now common practice for the production of combinatorial libraries, primarily in pharmaceutical companies, but is novel in the area of complex natural product synthesis as seen here. The key step is the C12–C13 bond-forming reaction of compound 6 orchestrated by the rutheni-

*Bristol-Myers Squibb has registered Taxol as a trademark and wishes the scientific community to use the name paclitaxel.

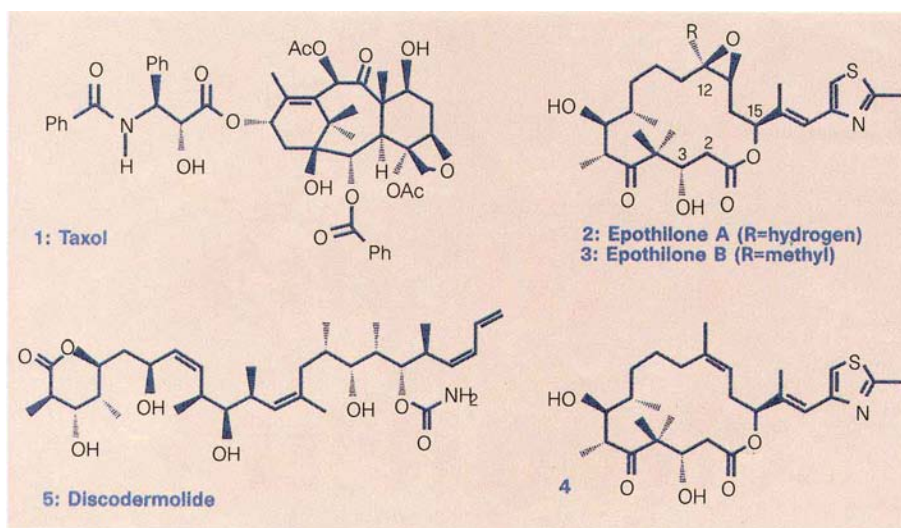


Figure 1 Some naturally occurring and synthetic cytotoxic agents that bind to microtubules.

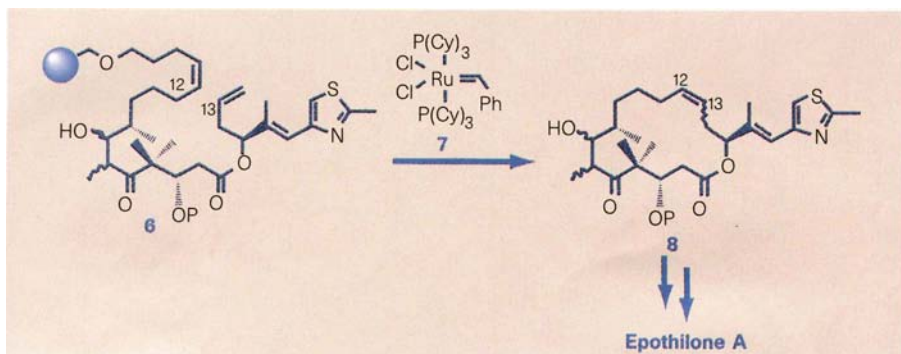


Figure 2 Olefin metathesis: a key step in the solid-phase synthesis of epothilone A, which in the process cleaves the linking chain to the resin bead (P = *t*-BuMe₂Si; Cy = cyclohexyl).

um catalyst 7 (ref. 8), which simultaneously generates the 16-membered macrocycle 8 and cleaves the polymer support. Although this process has been widely used in polymer synthesis, its potential in the construction of complex ring systems is only now being realized. Its application to the epothilone system has also been demonstrated independently by Danishefsky and Schinzer⁴.

The accompanying synthesis⁵ of epothilone B (3 in Fig. 1) is more traditional, in that a C15–O bond formation (a macrolactonization) is employed to form the large ring. In contrast, in the first reported synthesis⁹ of epothilone B, Danishefsky and co-workers used an unusual intramolecular aldol reaction at C2–C3 to close the macrocycle. Both these syntheses proceed via alkene 4 with a selective epoxidation as the final step.

Now that the synthesis of both epothilones has been achieved, chemists and biologists can begin to find out what features of the compounds are essential for their activity. Already, compound 4 has been found to be more active than either taxol or the epothilones in a tubulin-assembly assay, indicating that the epoxide may not be essential for microtubule stabilization^{5,9}. A library of further epothilone analogues can now be produced by varying the existing

synthetic routes^{4,5,9}, and those analogues with strong antiproliferative properties can go on to clinical evaluation. Unlike the early stages in the development of taxol³, which was originally extracted from the bark of the Pacific yew, work should not be hindered by the availability of these compounds. It is to be hoped that a promising epothilone analogue will retain its activity *in vivo*.

The unfinished epothilone story is yet another illustration of the wealth of biological activity associated with natural products, and the ingenuity of synthetic chemists in constructing complex molecules. As for whether the epothilones turn out to be better than taxol at fighting cancer, it is far too early to say. □

Cameron J. Cowden and Ian Paterson are at the University Chemical Laboratory, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK.

- Hofle, G. *et al.* *Angew. Chem. Int. Ed. Engl.* **35**, 1567–1569 (1996).
- Bollag, D. M. *et al.* *Cancer Res.* **55**, 2325–2333 (1995).
- Horwitz, S. B. *Nature* **367**, 593–594 (1994).
- Wessjohann, L. *Angew. Chem. Int. Ed. Engl.* **36**, 715–718 (1997).
- Nicolaou, K. C. *et al.* *Nature* **387**, 268–272 (1997).
- Haar, E. T. *et al.* *Biochemistry* **35**, 243–250 (1996).
- Hung, D. T., Chen, I. & Schreiber, S. L. *Chem. Biol.* **3**, 287–293 (1996).
- Grubbs, R. H., Miller, S. J. & Fu, G. C. *Acc. Chem. Res.* **28**, 446–452 (1995).
- Su, D.-S. *et al.* *Angew. Chem. Int. Ed. Engl.* **36**, 757–759 (1997).



100 YEARS AGO

Messrs. John Walker and Co., Ltd., have sent us several packets of their "Perfection Envelope," which have been designed to do away with the necessity of licking, or otherwise moistening, the gum on the flap of an envelope in order to make it adhere. The gum is placed upon the back of the envelope instead of on the flap, so that by moistening the flap the envelope can be sealed without touching the gum with the tongue. The advantage thus gained is not very clear, for it is possible in the case of any envelope to moisten the back instead of the gummed flap. What is really required is a simple device for sealing envelopes without going through the very objectionable practice of licking them.

From *Nature* 13 May 1897.

50 YEARS AGO

It has been known for a long time, particularly from the work of Schuster, Sutherland and H. A. Wilson, though lately little regarded, that the magnetic moment *P* and the angular momentum *U* of the earth and sun are nearly proportional.... We can write, in fact,

$$P = \beta \frac{G^{1/2}}{c} U,$$

where β is a constant of the order of unity.

For the first time, the magnetic field of a star, 78 Virginis, has recently been measured (Babcock, 1947). Using the best estimate available for its mass, radius and rotational velocity, the calculated value of the magnetic field agrees with its observed value. We have, therefore, a rough verification of the equation above for three bodies, the earth, sun and 78 Virginis, and covering a range of *P* and *U* of more than 10¹⁰:1, though only of 2,000:1 in measured field. It is therefore considered that the above equation must be taken seriously as a possible general law of Nature for all massive rotating bodies. ...

It is suggested tentatively that the balance of evidence is that the above equation represents some new and fundamental property of rotating matter. Perhaps this relation will provide the long-sought connexion between electromagnetic and gravitational phenomena.

Prof. P. M. S. Blackett, F. R. S.
From *Nature* 17 May 1947.

Simple rules with complex outcomes

Gareth Hughes and José Luis González Andújar

A crop is infested by weeds. Viewed in economic terms, the decision on whether or not to apply a herbicide has two components. The short-term component of weed-control decision-making is a comparison of the cost of control measures with the value of the reduction in crop yield that would occur in the current season if no action were taken. The weed population density that would cause a reduction in revenue equal to the cost of control measures is known as the economic threshold. But there is also a long-term component to decision-making, because weeds that are not controlled in the current season can contribute to the weed seedbank population in the soil, and so may lead to crop-yield reductions in future seasons.

Can we identify a threshold population density for the implementation of weed control that would be economically optimal over a number of seasons? Wallinga and van Oijen, writing in the current issue of *Crop Protection*¹, show in fact that the choice of threshold level does not affect the long-term frequency of need for weed control. This finding has implications for the use of thresholds as a basis for weed-control decision-making.

Although economic threshold is a

cornerstone of the theory of weed-control decision-making², it may be criticized as being, on the one hand, too complicated (from the point of view of practical implementation); or, on the other, as too simplistic (from the point of view of a systems analysis of the crop-weed agroecosystem). Those who view economic thresholds as too complicated can point to the amount of information required to calculate a threshold. The minimum data requirements are the potential yield of the crop and its likely value, the reduction in crop yield per unit weed population density, the efficacy of the treatment and its cost. We then need to estimate the weed population in the field, by sampling, to be able to compare the observed density with the calculated threshold, as a basis for decision-making.

Those who view economic thresholds as too simplistic can point out that this is a discrete-choice threshold: herbicide is either applied at a predetermined rate or not applied at all. The long-term component of weed control is omitted, and the calculation of the threshold takes no account of spatial pattern in the weed population, spatial variability in herbicide application, or the effects of an individual's decision on the

world beyond a particular weed-infested crop. There is also an implicit assumption that all of the factors involved in the calculation are known with certainty (or, alternatively, that the uncertainty attached to estimates of particular parameters does not affect the calculation).

It is possible to think of ways to modify threshold calculations to encompass greater realism. For example, Wiles *et al.*³ make the point that a description of the spatial pattern of a weed population amounts to information on the uncertainty attached to average weed population density. However, as Wallinga and van Oijen point out, the problem is not a lack of realism, but that the discrete-choice threshold may simply be an inappropriate basis for weed-control decision-making when the objective is to limit the frequency of herbicide use in the long term.

The starting point for a model of weed control^{4,5} is often a weed population dynamics model, to which a parameter is added to mimic the effect of control measures. Wallinga and van Oijen's model explicitly embodies the discontinuity in the threshold concept. Weed population density in the forthcoming season, N_{t+1} , is related to weed population density in the current season, N_t , by:

$$N_{t+1} = \begin{cases} aN_t & \text{if } N_t \leq K \\ bN_t & \text{if } N_t > K \end{cases}$$

in which K is the threshold population density for control, a is the per capita growth rate of the uncontrolled weed population ($a > 1$) and b is the per capita growth rate of the controlled population ($0 < b < 1$). This model, qualitatively suggested by data from studies⁶ of cereal crops infested by *Avena fatua* (wild oats), is known to exhibit quasiperiodic — 'chaotic' — temporal dynamics⁷. As industrial economist Art de Vany has said⁸, in a rather different context, "People follow simple rules, and this produces complex outcomes".

Although the weed population density in a particular season is not predictable, the fluctuations in weed population density are bounded. There is a lower bound when, in the previous season, the observed density was just greater than K and control measures were taken; and an upper bound when the observed density in the previous season equalled K and control measures were deemed unnecessary. A probability distribution of weed population densities within these bounds can be extracted, and Wallinga and van Oijen derive from this the average control frequency: the proportion of years in which control measures are necessary. This proportion depends on the parameters a and b , but — crucially — not on K . That is to say, the control frequency does not depend on the threshold level.

Is there an alternative to weed-control



Figure 1 Case for control? Patchy infestation of a winter-wheat crop with the weeds *Galium aparine* (cleavers) and some *Matricaria recutita* (wild chamomile). (Photograph courtesy of J. Wallinga.)

decision-making based on a discrete-choice threshold? Routine application of herbicides, without assessment of the requirement, is likely to be unacceptable on environmental grounds as well as being economically sub-optimal. The rate and timing of control measures could be matched to the level of infestation. This policy avoids problems associated with the use of a discrete-choice threshold and can be economically optimal⁹. But the data requirements for such an economically optimizing threshold are greater and, even then, the calculation is still open to the criticism of being oversimplistic.

Instead, Wallinga and van Oijen's analysis ought to concentrate attention on the need for preventive weed management¹⁰, involving the use of cultural and biological methods of weed control, supported by a concerted effort to reduce weed seedbank populations. At present, scientific understanding of preventive management is limited and largely qualitative. Increased understanding of two major issues is

required: the costs and risks associated with weed prevention need to be identified and quantified; and the underlying ecological mechanisms need to be characterized in order to determine the effect of preventive practices on the long-term dynamics of weed populations. □

Gareth Hughes is in the Institute of Ecology and Resource Management, University of Edinburgh, Edinburgh EH9 3JG, UK. José Luis González Andújar is in the Departamento de Protección de Cultivos, Instituto de Agricultura Sostenible, 14080 Córdoba, Spain.

1. Wallinga, J. & van Oijen, M. *Crop Protection* **16**, 273–278 (1997).
2. Auld, B. A., Menz, K. M. & Tisdell, C. A. *Weed Control Economics* (Academic, London, 1987).
3. Wiles, L. J., Gold, H. J. & Wilkerson, G. G. *Weed Res.* **33**, 241–252 (1993).
4. Doyle, C. J. *Crop Protection* **10**, 432–444 (1991).
5. Mortimer, A. M., Sutton, J. J. & Gould, P. *Weed Res.* **29**, 229–238 (1989).
6. Selman, M. *Proc. 10th Br. Weed Control Conf.* **3**, 1176–1188 (1970).
7. May, R. M. & Oster, G. F. *Am. Nat.* **110**, 573–599 (1973).
8. Cassidy, J. *New Yorker* **73**, 36–44 (1997).
9. Plant, R. E. *J. Econ. Entomol.* **79**, 1–6 (1986).
10. Jordan, N. J. *Production Agric.* **9**, 485–490 (1996).

RNA editing

Message change for a fat controller

James Scott

RNA editing describes the processes that change the information content of RNA from that encoded in the gene. The term encompasses a series of distinct traits with separate origins, although it does not include messenger-RNA splicing or hypermodification of bases in transfer RNA. But it can involve the insertion, deletion or modification of nucleotides and, on page 303 of this issue, Burns and colleagues¹ describe a study that expands the repertoire of this last form of RNA editing.

Trypanosomes manifest RNA editing in its most florid form. Mitochondrial RNA transcripts are altered by the deletion or addition of uridine (U) residues. The result is that abnormalities in local reading frames are corrected or, more spectacularly, extensive editing can create as much as half of the mRNA sequence². All higher plants must also edit many uridine-to-cytidine (C) transitions in RNA transcripts in their mitochondria and chloroplasts, to correct deviations from the universal genetic code³.

RNA editing in nuclear transcripts is much more discrete than in organelles³. A single codon in the mRNA for the mammalian *apolipoprotein-B* gene undergoes a C-to-U conversion. The mRNAs for certain glutamate-gated calcium channels and the hepatitis-delta virus RNA undergoes adenosine (A) to inosine (I) editing — inosine is

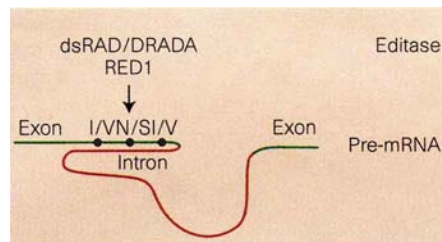


Figure 1 Burns *et al.*¹ have shown that transcripts encoding the serotonin (5-hydroxytryptamine, 5-HT) receptor subtype, 5-HT_{2C}, undergo RNA editing by the double-stranded RNA editing deaminases dsRAD/DRADA and RED1. Genomically encoded adenosine residues in the pre-mRNA are converted to inosines, resulting in the following amino-acid changes at three sites in the receptor: isoleucine (I) to valine (V); asparagine (N) to serine (S); and isoleucine to valine. The editing sites are identified in the double-stranded RNA that is formed between intron and exon sequences.

read as guanosine when it is translated by the ribosome.

The release of glutamate at synapses in the mammalian central nervous system activates cation-specific receptor channels (glutamate receptors; GluRs)^{3,4}. These include the α -amino-3-hydroxy-5-methylisoxazole-4-propionic acid (AMPA) receptors, and the

high-affinity kainate receptors. AMPA receptors mediate most of the fast excitatory neurotransmission at central synapses; kainate receptors are rapidly desensitized in the presence of kainate. The permeability of AMPA- and kainate-receptor channels can be dramatically altered by editing of the channel-subunit mRNAs at several distinct sites. These are the Q/R site — editing CAG (which encodes glutamine, Q) in the genomic sequence, to CGG (which encodes arginine, R) in the mRNA — in the GluR-B, GluR-5 and GluR-6 subunits; the R/G site in the GluR-B, GluR-C and GluR-D subunits; and the I/V and Y/C sites in the GluR-6 subunit. Each amino-acid change has important functional consequences for the permeability to calcium and gating of the channel. For example, transgenic mice that produce editing-incompetent GluR-B alleles⁵ develop epilepsy and die shortly after birth.

Editing sites are identified in the pre-mRNA (Fig. 1) by the formation of double-stranded RNA — sequences in nearby introns hybridize to complementary sequences that surround the editing site. Edited adenosines tend to occur in mismatched positions in loops and bulges, where the open geometry of the major groove of the RNA helix can allow the editing enzymes to gain access. Biochemically, the editing of A to I in the GluRs is a hydrolytic deamination⁶ that is carried out by double-stranded RNA adenosine deaminase (dsRAD or DRADA). dsRAD/DRADA strongly edits R/G sites of the GluR-B and the Q/R site of GluR-5 and GluR-6 transcripts, but it has a relatively low efficiency for other sites. There are also two new members of the double-stranded RNA adenosine deaminase family⁷, one of which — double-stranded RNA-specific editase (RED1) — is selective for the Q/R site in GluR-B pre-mRNA. As yet, the third member of the family, RED2, has no known substrate.

Burns *et al.*¹ have now expanded the targets for the double-stranded RNA adenosine deaminases to a series of sites found in the mRNA encoding a subtype of the serotonin (5-hydroxytryptamine, 5-HT) receptor, 5-HT_{2C}. Three amino-acid codons are edited from the genomic sequence so that, in the mRNA, the sequence coding for isoleucine (I) becomes valine (V) at the first site, asparagine (N) becomes serine (S) at the second, and I becomes V at the third (Fig. 1). The editing site-specificity in the 5-HT_{2C} receptor transcripts mimics that found in GluR-B transcripts, in that both dsRAD/DRADA and RED1 are necessary for full and efficient editing.

The altered residues lie in the second intracellular loop of the 5-HT_{2C} receptor. This is an important region for signal transduction, because it mediates coupling of the

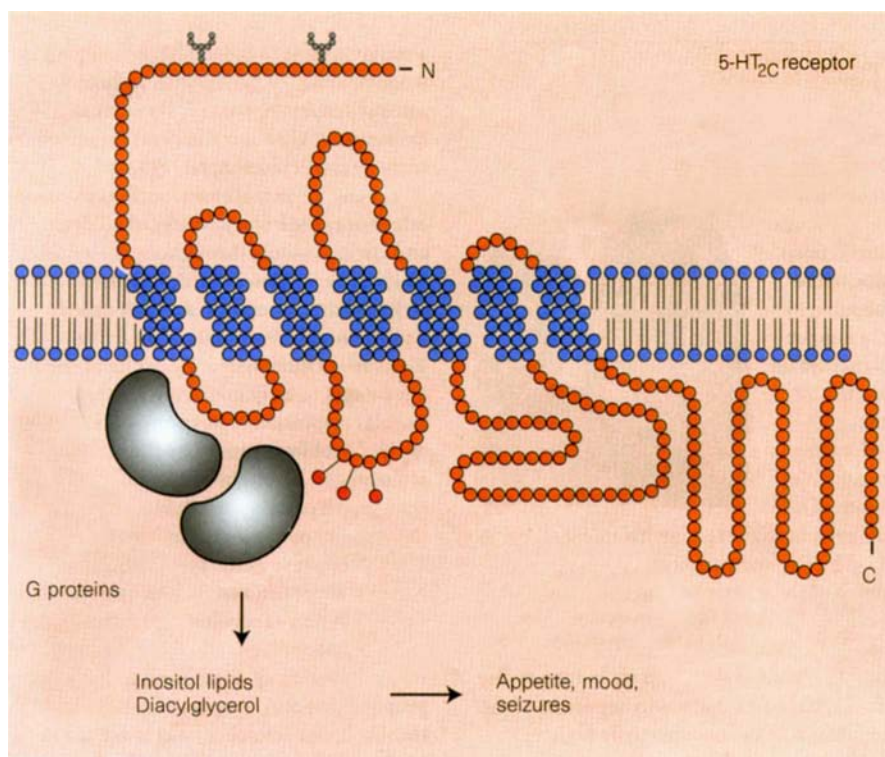


Figure 2 The amino-acid residues that are changed by RNA editing are confined to the second intracellular loop of the 5-HT_{2C} receptor. Edited receptors show decreased guanine-nucleotide-binding (G) protein coupling, which leads to a reduction in signalling by inositol lipids and diacylglycerol. Studies with 5-HT_{2C}-receptor knockout mice indicate that, in humans, this could result in decreased appetite, mood changes, and increased neuronal excitability which could lead to seizures. (Adapted from ref. 9.)

receptors to guanine-nucleotide-binding (G) proteins (Fig. 2). Receptors that are encoded by the fully edited transcript show a 10–15-fold reduction in G protein–receptor coupling, leading to decreased activation of phospholipase C (an enzyme that hydrolyses membrane phospholipids to activate the inositol phosphate and diacylglycerol signalling pathways). Transcripts that are not completely edited encode receptors that behave in the same way as unedited 5-HT_{2C}.

The 5-HT_{2C} receptors are widely expressed throughout the central nervous system, including the cortex, striatum, hypothalamus, olfactory bulb and choroid plexus. Regional differences in RNA editing are apparent. The most abundant edited variant found in most regions of the brain that were sampled by Burns *et al.* would not be predicted to be any less responsive to 5-HT than the unedited protein. Nevertheless, differential editing does suggest a new mechanism for the regulation of 5-HT-mediated signal transduction.

It is also interesting that mice in which the 5-HT_{2C} receptors have been inactivated by homologous recombination have an eating disorder and are overweight. Moreover, agonists of the 5-HT_{2C} receptors no longer act as appetite suppressants in these mice⁸. Intriguingly, drugs such as fluoxetine (Prozac) and dextfenfluramine, which

increase nervous transmission by blocking re-uptake of 5-HT, control the overeating that is associated with certain obesity syndromes, and act as antidepressants: these drugs most probably act through 5-HT_{2C} receptors. 5-HT_{2C}-receptor knockout mice die from epileptic seizures, and this can be mimicked in wild-type animals with non-selective 5-HT_{2C}-receptor antagonists. So the epilepsy is probably not due to abnormal brain development, but to the absence of a normal function of the 5-HT_{2C} receptor in mediating tonic inhibition of neuronal excitability. Perhaps RNA editing is involved in the control of higher cerebral functions such as appetite, mood and consciousness, or in diseases such as depression and epilepsy. And maybe RNA editing will now become a target for the pharmaceutical industry. □

James Scott is in the Department of Medicine, Royal Postgraduate Medical School, Hammersmith Hospital, Du Cane Road, London W12 0NN, UK.

1. Burns, C. M. *et al.* *Nature* **387**, 303–308 (1997).
2. Simpson, L. & Thiemann, O. H. *Cell* **81**, 837–840 (1995).
3. Scott, J. *Cell* **81**, 833–836 (1995).
4. Seeburg, P. H. *J. Neurochem.* **66**, 1–5 (1996).
5. Brusa, R. *et al.* *Science* **270**, 1677–1680 (1995).
6. Bass, B. L. *Trends Biochem.* (in the press).
7. Melcher, T. *et al.* *Nature* **379**, 460–464 (1996).
8. Tecott, L. H. *et al.* *Nature* **374**, 542–546 (1995).
9. Stryer, L. *Biochemistry* 3rd edn (Freeman, New York, 1988).

Daedalus

Reduced mass

Masonry is one of the earliest and most brilliant of inventions. Merely by piling rocks or bricks on top of one another, you can assemble a structure as basic as a hut or as complicated as a cathedral. All the joints are safely in compression, and are kept stable by simple friction. (The mortar between the stones is not an adhesive, but a load-spreader. With careful enough design, you can do without it, as proved by Egyptian pyramids and English dry stone walls.) There are two disadvantages. The whole thing is extremely heavy; and it can get very cold and damp in winter. Daedalus is now tackling both problems.

He points out that bricks and many stones are porous. This is not dangerous; to be stable, a masonry structure needs wall and pillars of such a large cross-section that the crushing loads on the bricks are extremely low. So he is taking this porosity to extremes. DREADCO ceramicists are inventing a light, foamed brick with little strength but equally little weight. They are mixing brick clay with a 'baking powder' whose grains give off carbon dioxide when heated to firing temperature. Cunningly, the powder is made of grains with a carefully adjusted set of sizes. The biggest grains blow enough bubbles to make a close-packed hexagonal lattice of bubbles in the clay. The next smallest blow bubbles to fill the interstices in this lattice; the smallest size blows bubbles in the interstices of the result. The final brick combines a rigid structure with an amazingly low density, and wonderful heat insulation.

To make good use of such bricks, they should be stacked together with a foamed mortar; so the ceramicists are inventing one. A fairly conventional mortar is mixed with detergent, foamed with compressed air, stored in a backpack and delivered by a sort of over-sized icing gun. Similarly, the interior walls and ceilings of the structure will be decorated with an equally light foamed plaster, and the roof will be covered with foamed tiles.

Most of the stresses in a normal building come from its own weight. So a foamed masonry building, much weaker but also much lighter than a conventional one, will be just as safe against gravitational collapse. The designer will merely have to secure it against wind loads and extreme internal distributions of people and things. It will be much easier to erect than a normal building, and much warmer when completed. And should an earthquake or a runaway vehicle bring it down, it will do far less damage.

David Jones

Lyman Spitzer Jr (1914–97)

Astrophysicist and founder of controlled nuclear fusion research

Lyman Spitzer, who died on 31 March, shaped four different fields of science. His theoretical work on star clusters, interstellar matter and laboratory plasmas is classic. He proposed, helped to design and nurtured the construction of the Hubble Space Telescope. And he proposed a method to control thermonuclear fusion for energy production in a magnetic field, and led a pioneering effort to do so at Princeton University.

Spitzer was born in Toledo, Ohio, the son of a businessman. At Yale he was drawn to physics, and went to Cambridge University to study astrophysics. Later, at Princeton, he studied with Henry Norris Russell, then the dean of American astronomy, and while a postdoctoral fellow at Harvard in 1938–39, he met the rising young astrophysicist Martin Schwarzschild* and they became lifelong close friends.

Spitzer undertook underwater sound research for the Navy in the Second World War, after which he returned briefly to Yale. It was there that he received the offer, in 1947, to succeed Russell at Princeton. Recruiting Schwarzschild as a colleague, he remained at Princeton as director of the observatory until 1979, and as professor until 1982.

In the 1940s, convinced that stars must be forming even today from the clouds of gas and dust that had been observed in interstellar space, Spitzer resolved to understand how. He conceived of an ultraviolet telescope in orbit, which could measure the spectra of atoms and molecules in interstellar space. In parallel, he worked on the theory of the heating and cooling of interstellar gases, and their response to the pressure of hot regions near luminous stars. His investigations established the study of interstellar matter as an independent and rich discipline, summarized in his books *Diffuse Matter in Space* (Wiley, New York, 1968) and *Physical Processes in the Interstellar Medium* (Wiley, New York, 1978).

His proposal for an orbiting telescope to study interstellar material was finally achieved with the launch of the Copernicus satellite in 1972. When a pre-launch test of the focusing mechanism revealed a failure in a crucial component, Spitzer saved the

mission by calculating — on the night of the scheduled launch — the focus corrections due to the bending of the support-structure on Earth and the refraction caused by a beam of hot air present in the test environment.

The Copernicus mission showed that much of the matter in space is in the previously unobserved form of hydrogen molecules. It also showed that most of the heavy elements are bound up in interstellar dust grains, and that one component of the medium had the unexpectedly high temperature of about a million degrees — unexpected, that is, by those who had not read Spitzer's prediction in 1956 of a hot gaseous corona of the galaxy, analogous to that of the Sun. Finally, deuterium atoms were observed for the first time in interstellar space, with an abundance that implies that we are living in an open Universe — unless there is a huge amount of dark matter in space.

Spitzer, together with H. Alfvén and a few other visionary pioneers, established the theoretical foundations of plasma physics in the 1950s. Recognizing the importance of the thermal, electrical and mechanical transport coefficients in a fully ionized gas, Spitzer made the initial calculations of their values, and also the first computations of containment, resistive heating and diffusion losses for a plasma confined in a torus. (The electrical conductivity of a fully ionized gas is still called the Spitzer conductivity.) This work was summarized in his book *Physics of Fully Ionized Gases* (Wiley, New York, 1962).

In 1951, Spitzer proposed to the Atomic Energy Commission the building of a reactor, which he called a 'stellarator', that would be "designed to obtain power from the thermonuclear reactions between deuterium and either deuterium or tritium". The idea came to him on a ski slope. Until 1967 he directed the resulting programme, Project Matterhorn (later the Princeton Plasma Physics Laboratory), and steady progress was made towards the goal of controlled thermonuclear fusion.

In stellar dynamics, Spitzer helped show how the process of 'relaxation' causes

a stellar system to inexorably approach a singular state — that is, with an infinite central density (Spitzer, L. *Dynamical Evolution of Globular Clusters* (Princeton Univ. Press, Princeton, NJ, 1987)).

Spitzer's seminal contributions to space astronomy are known to schoolchildren and curious adults throughout the world. In 1946, he proposed the development of large space telescopes in a report called *Astronomical Advantages of an Extra-Terrestrial Observatory*. He outlined the advantages to be gained from greater angular resolution (overcoming the 'seeing' problems caused by the turbulent atmosphere), from the increased wavelength coverage available, and from the stability of a low-gravity environment. These goals have guided generations of space scientists and engineers, and have led to a succession of revolutionary discoveries made in space.

The large space telescope that Spitzer proposed in 1946, eventually called the Hubble Space Telescope, was launched in 1990, with Spitzer and his family observing the event. In the years between 1946 and 1990, and indeed afterwards, Spitzer provided gentle but perceptive scientific, technical and political guidance that helped lead to the extraordinary successes of the Hubble Telescope. The modest and somewhat humorous way he regarded his own eminence is typified by a remark he once made to Bahcall in the 1970s during one of their innumerable trips in support of what was to become the Hubble Space Telescope: "You know, all the objectives of this trip would be fulfilled and much effort and money would be saved if they would just allow us, instead of appearing directly, to send wax images of ourselves."

Together with Martin Schwarzschild, Spitzer established in Princeton an enduring cordial atmosphere of mutual support and encouragement for astrophysical research at the highest level. On the day of his death, Spitzer worked a full day in his beloved Peyton Hall, talking enthusiastically to colleagues and making progress on a research paper.

He was admired and loved by all who knew him, non-scientists and scientists alike. He once confessed to one of us that, "I suppose I have at least one serious fault as a scientist: I love to work on big problems". But he not only worked on big problems, he solved them.

George B. Field & John N. Bahcall

George B. Field is at the Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. John N. Bahcall is in the Institute for Advanced Study, School of Natural Sciences, Princeton, New Jersey 08540, USA.

*Martin Schwarzschild died on 10 April. An obituary will appear here shortly.

Poultry virus infection in Antarctic penguins

Antarctic penguins appear to be relatively free of infectious diseases, although there is serological evidence of infection with a number of avian diseases found elsewhere¹⁻⁴. An infectious agent is suspected in one case of mass mortality in Adélie penguin chicks, *Pygoscelis adeliae*⁵ (Fig. 1), but there have been no confirmed reports of any major outbreaks of infectious disease. Here we report evidence for the occurrence of an avian pathogen, infectious bursal disease virus (IBDV), in wild Antarctic penguins. This raises concern for the conservation of avian wildlife in Antarctica.

IBDV is a pathogen of domestic chickens *Gallus domesticus*, but antibodies have been detected in a variety of wild aquatic bird species⁶. It affects lymphoid organs, primarily the bursa of Fabricius in chicks, the main site for the development of antibody-producing B-lymphocyte populations. Sub-clinical effects and immunosuppression, caused by even the less virulent strains of IBDV, retard growth and development and predispose the chick to opportunistic infections⁷. Morbidity and mortality rates in young chickens vary but can be high, especially in a newly emerged, globally spreading, high-virulence strain.

We inferred IBDV infection from the presence of specific antibodies in serum collected from both emperor (*Aptenodytes forsteri*) and Adélie penguins. We collected samples from 52 emperor penguin fledgling chicks (four to five months of age) from Auster Rookery (67° 23' S, 64° 02' E) in December 1995, and between September



Figure 1 Adélie penguin about to feed its chick.

1995 and February 1996 from 133 adult Adélie penguins from two colonies within 40 km of Mawson (67° 31' S, 62° 48' E) (Fig. 2). We used a standard virus neutralization test (VNT) to measure antibody titres⁸. Antibody titres of 1 in 80 or greater were regarded as positive⁹. Using this conservative criterion, the prevalence of positive VNT reactors was 65.4% in emperor penguin chicks and 2.1% and 2.6% in the two colonies of adult Adélie penguins. Infection with IBDV occurs principally in young birds¹⁰ and so prevalence of seroreactors would decline with age.

We found no antibodies in Adélie serum taken from either chicks ($n=17$) or adult birds ($n=26$) in January 1996 from a remote and rarely visited colony at Edmonson Point (74° 21' S, 165° 03' E) in the Ross Sea (Fig. 2). In a retrospective analysis, the prevalence of antibodies in adult Adélie penguin serum collected at Mawson in January 1991 was 1.5% ($n=136$).

IBDV is relatively resistant to inactivation

by heat, desiccation and chemical agents, and is contagious and highly infectious by the faecal-oral route¹¹. It is likely to remain an environmental contaminant of colonies between breeding seasons. The recent rapid spread of the new highly virulent strains throughout most of the poultry industry in the Northern Hemisphere is testament to its potential for widespread dissemination. A potent source of environmental contamination in Antarctica could be from careless or inappropriate disposal of poultry products, allowing access by scavenging birds such as the south polar skua, *Catharacta macrorhynchos*. Spread within Antarctica could be facilitated through the movement of people carrying the virus on contaminated footwear, clothing, equipment or vehicles.

Although clinical disease was not apparent in either species of penguin, further investigation is warranted. The presence of seroreactors near centres of human activity raises the possibility that the virus may have been introduced. The size of our chick sample at Edmonson Point was sufficient to detect a prevalence of in excess of 20% seroreactors ($P<0.05$)¹². If IBDV was present in this location, the seroprevalence would be expected to be similar to that among the equivalent age group of emperor chicks. Thus the absence of neutralizing antibodies in the chicks from this remote colony of Edmonson Point suggests that this area is free from IBDV contamination.

At this time, it must be assumed that IBDV may be pathogenic in Antarctic penguins, and is likely to be spread by human activity. The potential for expeditioners and tourists to be vectors of disease as they move around Antarctica may pose the greatest threat yet to its avian fauna.

Heather Gardner, Knowles Kerry Martin Riddle

Antarctic Division,
Department of Environment, Sport and Territories,
Channel Highway, Kingston 7050,
Tasmania, Australia
e-mail: heathe_gar@antdiv.gov.au

Sue Brouwer, Laurie Gleeson

Australian Animal Health Laboratory,
CSIRO, PO Bag 24, Geelong 3220, Australia

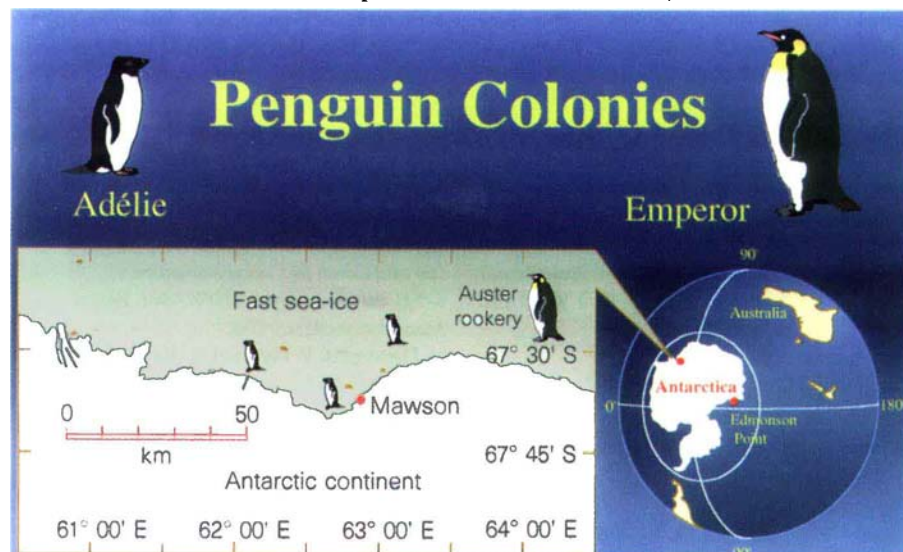


Figure 2 Antibodies to IBDV were identified in adult Adélie penguins at three sites within 40 km of Mawson and in emperor penguin chicks at Auster Rookery, 50 km east of Mawson. No antibodies were found in chicks or adult Adélie penguins from the remote colony at Edmonson Point in the Ross Sea.

- Clarke, J. R. & Kerry, K. R. *Korean J. Polar Res.* 4, 79-96 (1993).
- Morgan, I. R. & Westbury, H. A. *Avian Dis.* 25, 1019-1027 (1981).
- Morgan, I. R. & Westbury, H. A. *Hydrobiologia* 165, 262-269 (1988).
- Austin, F. J. & Webster, R. G. *J. Wildlife Dis.* 29, 568-571 (1993).
- Kerry, K. R., Gardner, H. G. & Clarke, J. R. *Microbiol. Aust.* 17, 16 (1996).
- Wilcox, G. E. *et al. Avian Pathol.* 12, 135-139 (1983).
- McFerran, J. B. in *Virus Infections of Birds* (ed. McFerran, J. B. & McNulty, M. S.) 213-228 (Elsevier Science, 1993).
- Manual of Standards for Diagnostic Tests and Vaccines* 496-503 (Office Internationale des Epizooties, 1996).
- Giambrone, J. J. *Avian Dis.* 24, 284-287 (1980).
- Hitchner, S. B. *Avian Dis.* 2, 611-613 (1976).
- Fenner, F. *et al. Veterinary Virology* (Academic, London, 1987).
- Cannon, R. M. & Roe, R. T. *Livestock Disease Surveys: A Field Manual for Veterinarians* (Australian Gov. Publ. Service, Canberra, 1982).

Infant leukaemia after the Chernobyl accident

In their Letter to *Nature*¹, Petridou *et al.* attributed an increased incidence of infant leukaemia in Greece to increased *in utero* exposure to ionizing radiation arising from the Chernobyl accident, which occurred on 26 April 1986. We see a similar increase in infant leukaemia in western Germany in children born after the Chernobyl accident. However, more detailed analyses of different contamination levels and dose rates show no relationship between exposure and incidence. We therefore conclude that the observed effect was not due to ionizing radiation from the Chernobyl accident.

Measurements of radioactive fallout and calculation of effective doses for the population show that large parts of Germany, especially in the south, experienced contamination comparable to that of Greece. The United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) estimated an effective dose-equivalent of 0.49 mSv attributable to the Chernobyl accident (over the following year) for the population in southern Germany and of 0.33 mSv for most of Greece².

There was a wide variation in levels of exposure in Germany, with UNSCEAR's first-year dose estimate for northern regions being 0.07 mSv, so it is possible to study potential dose-response patterns. Local fallout patterns and estimates of resulting average external and internal radiation doses are available for the 328 *Landkreise* of the former Federal Republic of Germany. Despite widespread long-distance transport of even fresh agricultural products in Germany, it has been shown that the variation in total dose closely follows ground-deposition patterns³, which thus can be used as a surrogate for dose. We chose levels above 10 kBq m⁻² ¹³⁷Cs to represent areas of high exposure and below 6 kBq m⁻² for low exposure⁴. Taking shielding but not iodine contribution (which adds little to blood stem-cell dose) into account, this translates into average *in utero* doses (total for nine months) of 75 and 55 µSv respectively, for the exposure levels defined above.

Since 1980, a population-based childhood cancer registry has existed, which receives reports of more than 90% of malignancies occurring in children up to 15 years of age, for the former Federal Republic of Germany⁵. Therefore, unlike the Greek study, we did not have to ascertain the number of cases of leukaemia retrospectively. Using the definitions of Petridou *et al.* for *in utero* exposure (children born between 1 July 1986 and 31 December 1987) there were 928,649 'exposed' children in the FRG (cohort B). Among these, 120,440 lived in the areas of high exposure. Cohorts A and C, defined as 'unexposed' groups, consisted of 3,601,176 children born between 1 January 1980 and 31 December 1985 (cohort A) and 2,029,613 children born between 1 January 1988 and 31 December 1990 (cohort C). The incidence of infant leukaemia in the 'exposed' cohort B is higher than in cohorts A and C (Table 1). The incidence rate did increase from cohort A (23.0 per million) to cohort C (29.6 per million), possibly due to some under-reporting in the initial phase of the German childhood cancer registry.

Looking at regions defined by radioactive ground deposition, the increase in leukaemia rate is highest in the regions with the lowest contamination by radioactive fallout. We believe that it is highly unlikely that this observation is due to misclassification, as there was a clear geographical distinction between the areas of high and low exposure.

In addition, because there were many short-lived radionuclides, and as surface contaminations in the human environment and on food were quickly washed out, radiation exposures from Chernobyl showed quite steep gradients with time. The dose rate in the first days after fallout were 23 times higher than in the last month of 1987. Any potential excess of infant leukaemia caused by intra-uterine radiation exposure should thus be most apparent in the older children of cohort B. If we divide B into two subcohorts (birth date July 1986 to March 1987 and April to December 1987) the rate ratio is 1.29 for the older subcohort and 1.67 for the younger subcohort (data for all areas, comparison with the combined cohorts A and C). These results again are not in accordance with the radiation hypothesis.

Analysing the incidence of all leukaemias that occurred in children up to 5 years old

shows that the relative increase in infant leukaemia seems to be compensated for in subsequent years, as the corresponding rate ratio of 'exposed' to 'unexposed' cohorts is 1.02 for the whole area (95% confidence interval, CI, 0.91–1.15). A decreased incidence in the second year of life was observed for the 'exposed' cohort (rate ratio, 0.84; 95% CI, 0.61–1.85).

Although our initial analysis is consistent with the observation of Petridou *et al.*, detailed trend analyses for contamination levels and critical time periods fail to correlate exposure levels with increased leukaemia rates. Thus we conclude that the observed increase of infant leukaemia is not caused by an increased *in utero* exposure to ionizing radiation from the Chernobyl accident.

J. Michaelis, U. Kaletsch

Institut für Medizinische Statistik und Dokumentation,
Johannes Gutenberg-Universität Mainz,
55101 Mainz, Germany

W. Burkart, B. Grosche

Institut für Strahlenhygiene,
Bundesamt für Strahlenschutz, Postfach 1108,
85758 Oberschleissheim, Germany

Petridou et al. reply — The increase in incidence of leukaemia among infants exposed *in utero* to radiation from the Chernobyl accident in Germany is strikingly similar to that noted in Greece. No explanation other than an *in utero* effect of ionizing radiation or a highly improbable chance manifestation can be seriously considered. In particular, no confounding influence with the required time-dependent properties can be invoked.

The lack of response to exposure in several subgroup analyses in Germany is hardly surprising, given the unavoidable non-differential exposure misclassification, the questionable correspondence between environmental measurements and personal exposures, and the sparse data. The apparent within-country response to exposure in Greece could be explained by the different cut-off points used in that study, the agro-economic practices in Greece that favour smaller-scale production-consumption systems or, conceivably, more benevolent chance.

E. Petridou, D. Trichopoulos

Harvard Center for Cancer Prevention,
677 Huntington Avenue, HSPH3-901, Boston,
Massachusetts 02115, USA

N. Dessypris, V. Flytzani, S. Haidas

M. Kalmanti, D. Kolioukas, H. Kosmidis

F. Piperopoulou, F. Tzortzatou

For addresses see ref. 1

Table 1 Cases of infant leukaemia

Region	'Exposed' birth cohort (B)			'Unexposed' birth cohorts (A+C)				
	Cohort size	Number	Incidence rate	Cohort size	Number	Incidence rate	Rate ratio	95% CI
Former FRG	928,649	35	37.7	5,630,789	143	25.4	1.48	1.02–2.15
Ground deposition (kBq m ⁻² ¹³⁷ Cs)								
<6	696,402	29	41.6	4,230,847	96	22.7	1.84	1.21–2.78
6–10	111,807	1	8.9	684,113	24	35.1	0.25	0.03–1.89
>10	120,440	5	41.5	715,829	23	32.1	1.29	0.49–3.40

Incidence, rates and rate ratios of children with acute leukaemia in the first year of life for the 'exposed' (B) and 'unexposed' (A+C) birth cohorts in the former Federal Republic of Germany and in regions with different levels of contamination. Incidence rate per 10⁶ children.

- Petridou, E. *et al.* *Nature* **382**, 352–353 (1996).
- United Nations Scientific Committee on the Effects of Atomic Radiation (United Nations, New York, 1988).
- Bayer, A. *et al.* ISH-Bericht 140/89 (Institut für Strahlenhygiene, Oberschleissheim, 1989).
- Michaelis, J. *et al.* *Klin. Pädiatr.* **208**, 172–178 (1996).
- Kaatsch, P., Haaf, H. G. & Michaelis, J. *Eur. J. Cancer* **31**, 993–999 (1995).

Animal mitochondrial DNA recombination

Genetic recombination is known to be a source of mitochondrial DNA (mtDNA) variability in plants, fungi and protists¹, but there continues to be a consensus (based on studies of somatic cell hybridization² and DNA repair³) that such processes do not operate on animal mtDNA⁴. Contrary to this opinion we have now identified and characterized the end-products of recombination in the mitochondrial genome of the phytoneurmatode *Meloidogyne javanica*.

Changes in the copy number of variable number tandem repeat (VNTR) sequences in mitochondrial DNA have occasionally been suggested to occur by recombination⁵⁻⁸, although other mechanisms could explain such modifications⁴. If a locus consisted of tandem repeats (frequently found in animal mtDNA⁹) interspersed with unique sequences, it would be possible to discriminate recombination from mechanisms that simply change the repeat copy number. Proximity to an origin of replication could also favour the detection of recombination, as displacement loops are associated with mtDNA replication⁹ and many models of recombination¹⁰.

The repeat region of *M. javanica* mtDNA (Fig. 1a) meets these criteria and provides an excellent tool with which to detect recombination in animal mtDNA. We analysed the sequences of these nematode VNTRs and frequently found deletions in the centre of the array (Fig. 1b), indicating that genetic recombination might be occurring.

A defining feature of recombination models is the breakage and rejoining of participating DNA strands. The products of intramolecular recombination in mtDNA would be reciprocal, double-stranded subgenomic circles closed by covalent bonds. One such subgenomic circle (a 'minicircle'; Fig. 1c) would contain only the region excised from the mitochondrial genome, whereas its reciprocal partner (a major circle; Fig. 1b) would retain the remainder of the wild-type mtDNA. Mitochondrial genomes containing deletions, similar to major circles, have been observed in several organisms¹¹, but the presence of subgenomic minicircular end-products (Fig. 1c) has, to our knowledge, not been demonstrated. The existence of these molecules would provide unambiguous support for recombination within animal mtDNA. We have now detected such small circular molecules composed of *M. javanica* mtDNA sequences using the polymerase chain reaction (PCR; Fig. 1c-e) and subsequent sequence analysis.

If recombination occurred, the mini-

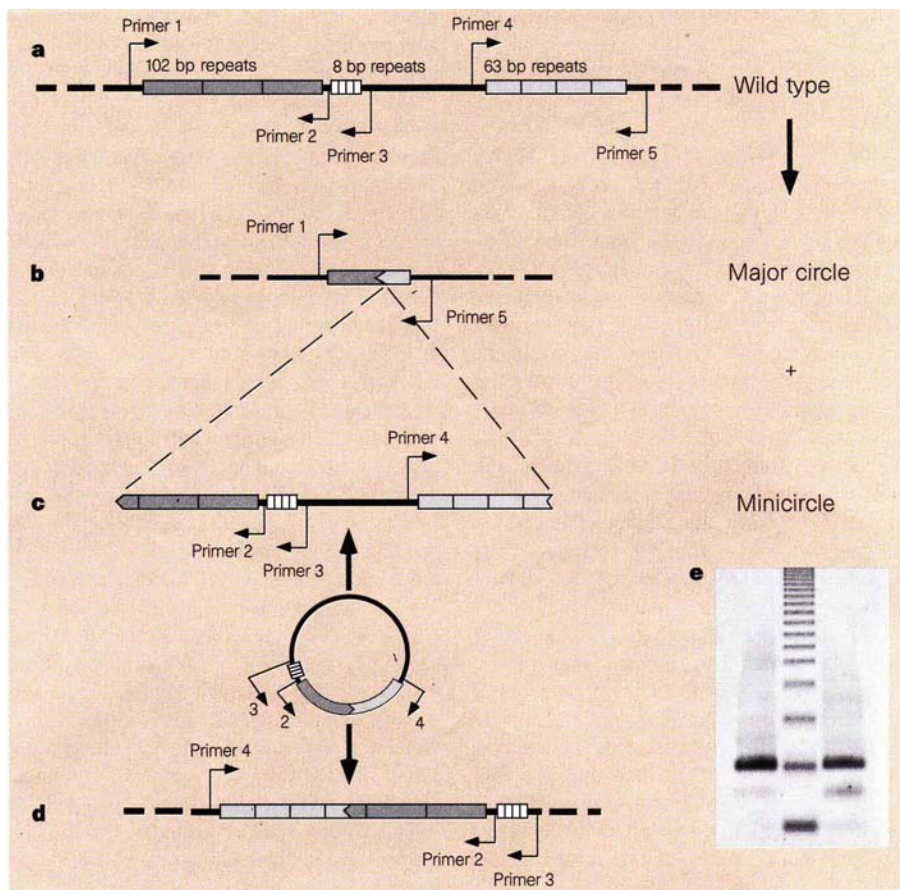


Figure 1 a, Organization of the *M. javanica* wild-type mtDNA¹⁴ major non-coding region. PCR primer sites are indicated by arrows, unique flanking sequences by solid black bars and repeat sequences by variously shaded rectangles. The sequence extends beyond the region depicted, and joins to form a circle. b, Major circle and c, minicircle recombination end-products. d, Primers 2 and 4 are divergent in orientation on the wild-type mtDNA and will only amplify a subgenomic product if the excised fragment forms a minicircle. e, Amplified products from this circularity assay (lanes 1 and 3) with a 123-base-pair standard (lane 2). The main products are ~250 base pairs (bp). Sequence analysis of cloned PCR products confirmed that they contained the predicted repeat sequence organizations. Sequences are available on request from the authors.

circles would have unique junctions that could not be produced by a deletion process that did not involve circularization of the DNA (Fig. 1d). Sequence analysis of eight cloned minicircle PCR products revealed that, after excision, the 3' end of the 63-base-pair repeat array merges with the 5' end of the 102-base-pair repeat array. Moreover, the nucleotide sequences of the two end-products are of the predicted reciprocal arrangement (Fig. 1b,d).

There are many described examples of VNTRs in animal mtDNA⁹. Changes in copy number and the maintenance of heteroplasmy have frequently been attributed to slipped-strand mispairing or illegitimate elongation¹². The nature of the sequence changes predicted by these models is not consistent with our data.

We have shown that the generation of variability in sequence organization in *M. javanica* mtDNA is entirely consistent with recombination. Sequence polymorphisms and deletions in mtDNA are implicated in many human pathologies¹³, so identification of the mechanisms responsible, their

frequency and their epidemiology is imperative. Recombination in the mtDNA of higher metazoa remains undocumented, but as it has now been observed in most eukaryotes, its presence seems highly likely.

David H. Lunt, Bradley C. Hyman

Department of Biology, University of California, Riverside, California 92521, USA
e-mail: bhyman@ucr.edu

- Gray, M. W. *Annu. Rev. Cell Biol.* **5**, 25-50 (1989).
- Hayashi, J., Tagashira, Y. & Yoshida, M. *C. Exp. Cell. Res.* **160**, 387-395 (1985).
- Clayton, D. A., Doda, J. N. & Friedberg, E. C. *Proc. Natl. Acad. Sci. USA* **71**, 2777-2781 (1974).
- Moritz, C., Dowling, T. E. & Brown, W. M. *Annu. Rev. Ecol. Syst.* **18**, 269-292 (1987).
- Rand, D. M. & Harrison, R. G. *Genetics* **121**, 551-569 (1989).
- Baumer, A., Zhang, C., Linnane, A. W. & Nagley, P. *Am. J. Hum. Genet.* **54**, 618-630 (1994).
- Zhang, D. X., Szymura, J. M. & Hewitt, G. M. *J. Mol. Evol.* **40**, 382-391 (1995).
- LaRoche, J., Snyder, M., Cook, D. I., Fuller, K. & Zouros, E. *Mol. Biol. Evol.* **7**, 45-64 (1990).
- Wolstenholme, D. R. *Int. Rev. Cytol.* **141**, 173-216 (1992).
- Dressler, D. & Potter, H. *Annu. Rev. Biochem.* **51**, 727-761 (1982).
- Hyman, B. C. & Slater, T. M. *Genetics* **124**, 845-853 (1990).
- Buroker, N. E. *et al. Genetics* **124**, 157-163 (1990).
- Wallace, D. C. *et al. Biochim. Biophys. Acta* **1271**, 141-151 (1995).
- Okimoto, R., Chamberlin, H. M., MacFarlane, J. L. & Wolstenholme, D. R. *Nucleic Acids Res.* **19**, 1619-1626 (1991).

Collecting interstellar dust grains

Interstellar dust grains are swept into the Solar System by the motion of the Sun through the interstellar medium. The velocity of interstellar grains as they enter the Earth's atmosphere is modulated by the Earth's motion around the Sun, reaching a minimum when the Earth is moving in the opposite direction to the Sun. Modelling of the entry heating in that season indicates that some of the interstellar dust particles enter the Earth's atmosphere without vaporizing, and that they could be retrieved from the Earth's stratosphere in the same way as the interplanetary dust particles that are collected by NASA aircraft.

Ion trails of dust particles have been detected¹ entering the Earth's atmosphere at velocities that significantly exceed the Solar System escape velocity, identifying these particles as interstellar. There is a seasonal variation in the flux, the peak flux probably indicating the time of year when the Earth, in its heliocentric orbit, is moving towards a stream of interstellar particles. The direction of movement increases their geocentric velocity so that particles as small as 15 μm produce detectable meteors¹. Six months later, when the Earth is moving away from the stream, the ecliptic component of the particle's geocentric velocity is reduced by twice the Earth's heliocentric velocity of 30 km s^{-1} , so that at the two extremes the atmospheric entry velocities are given by:

$$v_{\text{entry}} = ((v \sin i)^2 + (v \cos i \pm 30)^2)^{1/2} \quad (1)$$

where i is the inclination angle, and v is the heliocentric velocity of the particle at the time of entry.

Taylor *et al.*¹ indicate that most of the interstellar dust is in a stream (their stream A), which has an atmospheric entry velocity of about 90 km s^{-1} on day 32, the peak of the flux, and which is inclined at roughly 30° to the ecliptic plane. Six

months later, these particles enter the Earth's atmosphere at a velocity of approximately 39.6 km s^{-1} .

I modelled the peak temperature reached by a dust particle entering the Earth's atmosphere using the methods described in refs 2 and 3. This correctly predicted that interplanetary dust particles would survive atmospheric entry. Interstellar grains of diameter 15–40 μm have been detected striking the Earth¹, so I calculated the distribution of peak temperatures for 15 μm grains of four densities (0.5, 1.0, 2.0 and 4.0 g cm^{-3}), spanning the range from porous interstellar aggregates of silicate and carbonaceous material⁴ to compact circumstellar grains extracted from meteorites⁵ (Table 1).

For an entry velocity of 39.6 km s^{-1} , all of the 15- μm -diameter particles of density 0.5 g cm^{-3} , 21% of the particles of density 1 g cm^{-3} , and 5% of the particles of density 2 g cm^{-3} reach peak temperatures below 1,900 K—the melting point of typical extraterrestrial silicates. Although interstellar silicates smaller than 15 μm in diameter would not produce detectable radar signals, the observation of interstellar grains of diameter 15–45 μm (ref. 1) indicates that smaller grains might also be present. If so, these grains would be heated less severely than the 15 μm grains modelled, and grains in the 5–10 μm size range would have a significantly higher probability of surviving entry unaltered.

A stream of 0.5- μm -diameter interstellar grains, with a velocity of approximately 26 km s^{-1} , was detected at a distance of roughly 5.0 astronomical units (AU) from the Sun by the Ulysses spacecraft⁶. Modelling indicates that, at some times in the solar cycle, these particles can reach the Earth⁷. Assuming the same 30° inclination to the ecliptic, these 0.5 μm interstellar grains should enter the Earth's atmosphere at roughly 29 km s^{-1} when the Earth is moving away from the stream.

To model the peak temperature reached by 0.5 μm grains on atmospheric entry requires an assumption of the thermal emissivity. It has been shown⁸ that the

emissivity for olivine spheres declines from near one, as assumed in the Whipple model², to approximately 0.2 as the size of the particle decreases to 5 μm . Assuming a lower emissivity of 0.1 for the 0.5 μm interstellar grains, the peak temperatures reached are much lower than for the 15 μm grains (see Table 1).

Thus, significant fractions of both the large interstellar grains detected by Taylor *et al.*¹ and the smaller grains detected by the Ulysses spacecraft⁶ are expected to survive atmospheric entry when the Earth is moving away from the interstellar dust stream. During this season, the interstellar grains will settle into the Earth's stratosphere in the same manner as the interplanetary dust particles routinely collected by NASA sampling aircraft. Interstellar particles in the size range 5–20 μm may be present on NASA collection surfaces that are already flown, but redesigned collectors would be required for the efficient collection of the 0.5 μm particles.

Convincing identification of an individual particle as interstellar would come from detection of non-solar isotopic compositions of either the refractory elements (for example oxygen or silicon) or the trapped noble gases, both of which have been used to identify interstellar grains from meteorites⁵. However, some interstellar grains are likely to have solar isotopic ratios. These might be identified by their extreme radiation damage⁹ or by observing the seasonal variation in particles having chemical compositions and structures consistent with those inferred for interstellar dust. This would provide laboratory samples of the most abundant dust of the interstellar medium rather than just the atypical grains isolated from meteorites.

G. J. Flynn

Department of Physics,
State University of New York,
101 Broad Street, Plattsburgh,
New York 12901-2681, USA

Table 1 Peak temperature distribution of interstellar dust particles

Particle		Fraction of incident particles not heated above temperature						
Diameter (μm)	Density (g cm^{-3})	800 K	1,000 K	1,200 K	1,400 K	1,600 K	1,800 K	1,900 K
15.0	0.5	8.9×10^{-4}	5.2×10^{-3}	2.2×10^{-2}	7.6×10^{-2}	2.2×10^{-1}	5.7×10^{-1}	1.0
15.0	1.0	2.1×10^{-4}	1.2×10^{-3}	5.3×10^{-3}	1.8×10^{-2}	5.3×10^{-2}	1.4×10^{-1}	2.1×10^{-1}
15.0	2.0	5.2×10^{-5}	3.1×10^{-4}	1.3×10^{-3}	4.6×10^{-3}	1.3×10^{-2}	3.4×10^{-2}	5.2×10^{-2}
15.0	4.0	1.3×10^{-5}	7.7×10^{-5}	3.3×10^{-4}	1.1×10^{-3}	3.5×10^{-3}	8.9×10^{-3}	1.4×10^{-2}
0.5	1.0	1.2×10^{-2}	7.5×10^{-2}	3.2×10^{-1}	1.0	1.0	1.0	1.0
0.5	3.0	1.3×10^{-3}	8.1×10^{-3}	3.4×10^{-2}	1.2×10^{-1}	3.4×10^{-1}	8.9×10^{-1}	1.0

Entry velocity 39.6 km s^{-1} for 15 μm particles, 29.0 km s^{-1} for 0.5 μm particles.

1. Taylor, A. D. *et al.* *Nature* **380**, 323–325 (1996).
2. Whipple, E. L. *Proc. Natl Acad. Sci. USA* **36**, 687–695 (1950).
3. Fraundorf, P. *Geophys. Res. Lett.* **19**, 765–768 (1980).
4. Greenberg, J. M. & Hage, J. I. *Astron. Astrophys.* **286**, 915–924 (1994).
5. Zinner, E. *Science* **271**, 41–42 (1996).
6. Grün, E. *et al.* *Astron. Astrophys.* **286**, 915–924 (1994).
7. Gustafson, B. A. S. & Lederer, S. M. in *Physics, Chemistry and Dynamics of Interplanetary Dust*, ASP Conf. Ser. **V104**, 35–38 (Astron. Soc. Pacific, San Francisco, 1996).
8. Rizk, B. *et al.* *J. Geophys. Res.* **96**, 1303–1314 (1991).
9. Bradley, J. P. *Science* **265**, 925–929 (1994).

scientific correspondence is intended to provide a forum in which readers may raise points of a scientific character. Priority will be given to letters of fewer than 500 words and 10 references.

Still on the learning curve

The Language of Shape: The Role of Curvature in Condensed Matter – Physics, Chemistry and Biology

by Stephen Hyde, Sten Andersson, Kåre Larsson, Zoltan Blum, Tomas Landh, Sven Lidin and Barry W. Ninham

Elsevier: 1997. Pp. 383, Dfl370, \$231.25

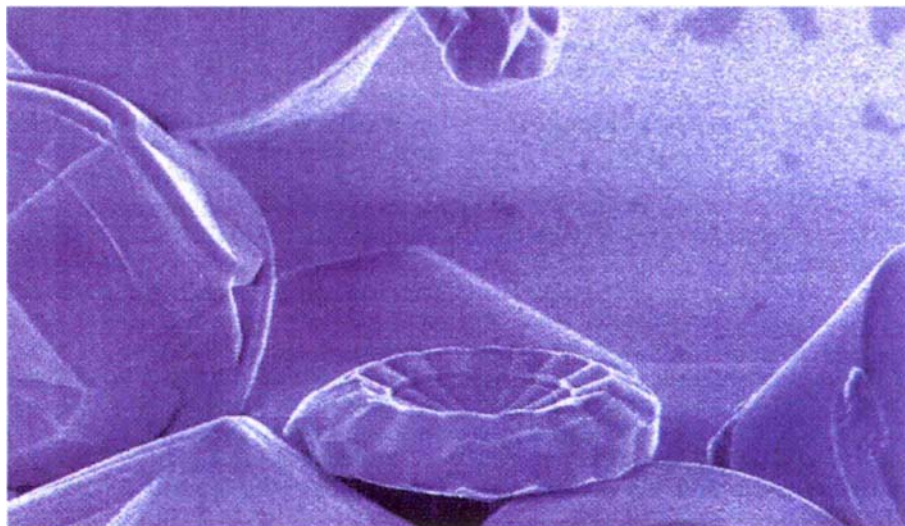
Alan L. Mackay

A mixture of cyclohexane and aniline, when above room temperature, forms a single phase, but on cooling separates into two layers. On warming up, fluctuations in the interface increase in magnitude and wavelength, and the interface then vanishes at the critical temperature where the single phase becomes stable. Perhaps interfaces between scientific subjects are like that; *The Language of Shape* is an interface book, rather knobbly and uneven in its contents, but anticipating the coalescence of regions of physics, chemistry and biology. Although there is collective authorship by the seven writers, individual contributions are apparent, like large fluctuations in refractive index. They ask us to see surfaces rather than molecules.

The authors belong to groups in Canberra, Australia, and Lund, Sweden, which have been working to bring new mathematical and geometrical insights into chemistry with the aid of modern computer graphics and numerical calculations. Traditional chemistry has concentrated on the behaviour of isolated molecules in solution and has moved on to long chains and surface chemistry, and thence to solid-state behaviour. Following Linus Pauling, the three-dimensional structure of materials has usually been seen in terms of linked coordination polyhedra, but Stephen Hyde and his coauthors move up a level in the hierarchy of structure and emphasize the surfaces, interfaces and equipotentials produced by assemblies of molecules, where the salient variable is curvature. Curvature appears most clearly in dealing with sheets of molecules which can take up a great variety of topologies, many of them periodic in three dimensions.

Just as complex mathematical structures can arise from simple generating formulae — the Mandelbrot set is generated from $z' = z^2 + c$ — so elaborate surfaces can appear from, for example, the requirement that the local mean curvature (the divergence of the unit vector normal to the surface) should be zero, plus boundary conditions. Complex long-range chemical structures may similarly arise from the requirements of simple local interactions.

Any such book must be compared with that published in 1917 by D'Arcy Thompson, *On Growth and Form* (Cambridge University Press, 1942, 2nd edn), which has long



Complex forms due to spontaneous curvature developed during the precipitation of mesoporous silica.

demand a sequel because Thompson hardly took note of atomicity. This new volume is not it, but it is a reconnaissance in strength of what may eventually appear. In 1936 Joseph Needham pointed out that “there are whole branches of mathematics dealing with the assessment of complexity, which no one has the ability or the imagination to make use of for grasping the biological situation” (*Order and Life*; MIT Press, 1968). In this latest book we have mainly the application of topology and the differential geometry of surfaces.

The book works through the basic mathematics of surfaces, then moves up from the atoms to the conceptual surfaces on which atoms can be seen to be concentrated, then to the forces of self-assembly and to the geometry of self-assembled systems to be found in liquid crystals, where individual molecules produce statistically anisotropic liquids. Copolymers may have very complex interfaces between the two phases. Lipids may form oriented monolayers and bilayers, leading to sheets and vesicles and ultimately to cell membranes. Crystals appear outmoded in comparison to the ingenuity of living systems, but periodicity sometimes reappears at higher levels of aggregation.

The longest chapter, based largely on the work of Tomas Landh, scours the literature for pictures of periodically structured cell membranes (‘cubic membranes’). With computer reconstructions it convincingly explains hundreds of electron micrographs in terms of continuous sheets convoluted into the three-dimensionally periodic minimal surfaces, rather than as isolated units ordered in arrays. It would be hard to imagine the three-dimensional structure of a cabbage from a few two-dimensional sections, and this is also the case for the endoplasmic

reticulum. We are everywhere encouraged to look anew at familiar structures and to see, for example, equipotential surfaces rather than discrete atoms (although at the level of single atoms Hyde’s explanation of the martensitic transformation in steel remains an intriguing speculation).

Even more elaborate molecular assemblies are considered, and suggestions made for new attitudes to specific assemblies: protein molecules, starch grains, muscle and collagen. The authors are missionaries and seek to convert us to their ways of looking at familiar structures. Morphogenesis is today an extremely active topic, and intimations of a junction between synthetic chemistry and biology are appearing with the generation of frameworks resembling the shells of diatoms, where amorphous hydrated silica frameworks can be generated in the interstices of periodic arrays of large organic molecules. The organic material is then burnt out, leaving a three-dimensionally periodic lattice of silica with tunnels of width around 100 Å.

This work seeks to reinterpret the geometry of structures in new terms, asking us to see sheets rather than molecules. For this it is essential to re-educate one’s spatial imagination with pictures, many of which are provided, but which represent intrinsically three-dimensional structures for which one really needs to get off flat pages and make solid models. Moving up from the micro-world of atoms, this new outlook may yet meet, in the meso-world of the biological structures revealed by the electron microscope, that of D’Arcy Thompson moving down from the macro-world of the naturalist. □

Alan L. Mackay is in the Department of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK.

Questions for our fairy godmother

Why Aren't Black Holes Black? The Unanswered Questions at the Frontiers of Science

by Robert M. Hazen and Maxine Singer

Anchor: 1997. Pp. 309. \$12.95

Lewis Wolpert

Being a scientist gives Stephen Jay Gould a continuing sense of privilege and reward. Science is for him the best technique we have for gaining knowledge about the world. He sees relativism — the argument that science is merely a social construct — as “perniciously wrongheaded”, and dismisses claims that science is finished as being astonishingly superficial. We remain ignorant about many of the deepest questions, and this is precisely what this book — which contains an introduction by Gould — aims to illuminate.

Robert M. Hazen and Maxine Singer try to examine the limits to which we have probed the physical and biological world. This is a grand ambition and the result is a mixed success. Starting with an analysis of the nature of scientific questions, they recognize that some questions are more fundamental than others. Peter Medawar coined a profound aphorism when he spoke of science as being the art of the soluble. Asking the right questions is at the very heart of great science and is probably more difficult than actually solving them. I fantasize about Goofgoos, the Good Fairy Godmother of Science, to whom one could put just one question. Every scientist should know exactly what question they would ask, noting that only a very brief answer will be given. It was just such questions that I hoped this book would spell out. But much of the book is instead a well-written popular review of

the whole of science — well, nearly all, because psychology, economics and sociology are absent.

The authors start with cosmology, the age and future of the Universe and why the Universe is lumpy, but also ask some questions that seem just outside science, such as whether life has meaning. They ask where are the WIMPs, the weakly interacting massive particles, that might provide the answer to all that embarrassing missing dark matter — 99 per cent of the Universe is just nowhere to be seen. Goofgoos, help please. She could also help with the Theory of Everything. But here there remains a problem that the authors do not explore. For, even if a cohesive set of equations is discovered from which everything else can be derived, we will still always be left with that seemingly unanswered question as to what explains that theory.

When discussing chemistry in terms of how atoms combine, Hazen and Singer suggest that, while molecules provide the basis of life, it may never be possible to predict the behaviour of living things from fundamental principles alone. This is an important question which relates to a reductionist issue that pervades the whole book. What do we regard as a satisfactory explanation for any set of phenomena? Biology has been brilliantly successful in explaining cell behaviour and inheritance in molecular terms, but is going down to the next level necessary or even interesting? Do we have to know the fundamental chemistry of enzyme action to explain metabolism or DNA replication?

Nowhere is this more evident than in relation to human and animal behaviour. In what terms are we to explain them? What would we regard, for example, as a satisfactory account of love or depression? It is easier to think about the physical basis of memory in terms of the synapses and molecular changes that the authors discuss, but we still have a very long way to go.

Topics covered include the quest for energy, population growth, the origins of the Earth and life itself. It is a fun if occasionally irritating book to read; the authors have, for example, no real insight into my own subject, developmental biology, and their evolutionary biology is weakish. There are also repeated references to the so-called ethical issues raised by science. But I learned a great deal about topics such as the inner structure of the Earth and the serious attempts to find life elsewhere in the Universe. The book gives in an accessible form a nice feel of what science is about. □

Lewis Wolpert is in the Department of Anatomy, University College London, Medawar Building, Gower Street, London WC1E 6BT, UK.

The chemist as accountant

Lavoisier: Chemist, Biologist, Economist

by Jean-Pierre Poirier

University of Pennsylvania Press: 1997.

Pp. 516. \$49.95, £47.50

Maurice Crosland

The work of Antoine Lavoisier (1743–1794) marked a turning point in the history of chemistry. Following the ‘chemical revolution’ of the late eighteenth century, of which he was the central figure, chemistry emerged almost as a new science. Indeed, it became *the* model science at a time when physics still lacked unity and coherence. Lavoisier not only reordered much of the science around oxygen, he also drew up a list of elements as the building blocks of inorganic chemistry. Only on the basis of these elements was it possible a generation later for John Dalton to introduce his atomic theory.

It would be difficult for a modern scientist to understand a chemistry book written before the time of Lavoisier, since in 1787 the leading French chemists collaborated to replace the previously chaotic nomenclature by a simple rational system, in which the names of compounds reflected their constituents.

Most readers would therefore expect a book on Lavoisier to be about his chemistry. Although Jean-Pierre Poirier tells the chemical story, he also tells much more, since his original interest lay in the financial and economic aspects of Lavoisier's life. For Lavoisier was very much more than a chemist.

One must not fall into the trap of supposing that all chemists started with a scientific education and then used their chemical expertise as the basis of their livelihoods. Lavoisier was educated in the classics, followed by a degree in law. He did, however, attend one course of chemistry

New in paperback

Blondes in Venetian Paintings, the Nine-Banded Armadillo, and Other Essays in Biochemistry

by Konrad Bloch

Yale University Press, \$16, £11

“The author displays a wide appreciation of metabolism, nutrition and comparative biology. The essays vary greatly in originality and new insights, although they all have something of interest for a diverse readership”, wrote Wilfred Niels Arnold in *Nature* 374, 130 (1995).

Driving Force: The Natural Magic of Magnets

by James D. Livingston

Harvard University Press, \$14.95, £9.95

“A smashing book.... In an era of

unfortunate public distrust and misunderstanding of the scientific method, I wish there were more science books for the general population written with so engaging an approach”, wrote Paul M. Grant in *Nature* 380, 679 (1996).

The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA

by Diane Vaughan

University of Chicago Press, \$14.95

“[Vaughan] has delved deep into the archives and conducted interviews with many of the protagonists”, wrote Harry Collins in *Nature* 380, 297 (1996).

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Lavoisier's involvement in taxation led to his arrest — when he went to the guillotine, it was not because of his scientific work, but despite it.

lectures. As regards his career, he could never have claimed chemistry as his occupation. For him it was a hobby, albeit one pursued with great zeal. His occupation was that of inspector of indirect taxes in a privatized system under the *ancien régime* in France where, having become a shareholder, he performed certain duties, and the tax firm — or 'farm' as it was called — received a proportion of the taxes collected as profit.

Lavoisier was therefore well schooled in finance. He was also an excellent administrator and served on many government commissions, including studies of agriculture, population, the Paris hospitals, slaughterhouses, mesmerism, the metric system and many other topics. He was also enrolled by the government to serve on the Gunpowder Commission, where he did sterling work.

Many readers might consider Lavoisier's financial work as irrelevant to his science, but there is evidence that he brought to chemistry something of the mentality of the accountant. He drew up a balance sheet in his study of chemical reactions, where the total weight of the products must be equal to the weights of the reactants. He adopted the principle of conservation of matter as a fundamental axiom and introduced some primitive chemical equations.

The centre of Lavoisier's scientific life was the prestigious Paris Academy of Sciences. Unlike the eighteenth-century Royal Society of London, membership of the academy was strictly limited and it showed Lavoisier's early promise that he was elected to it at the age of 24. We see how he became increasingly dissatisfied with the theory of phlogiston and slowly built up an alternative oxygen theory. In the original French edition of this book, the oxygen theory was presented as no more than a theory of combustion. Reviewers of that edition reminded Poirier of more recent research which, for example, pointed out that it was as much a theory of acidity and composition. Such criticisms have been gratefully taken on board by the author, so that this is more than a simple translation, it is a substantially revised edition. Poirier is generous in his acknowledgement to other spe-

cialists and provides an extensive bibliography. The book can therefore be recommended as a good biography which brings together much recent research.

This is certainly the definitive study of Lavoisier as financier. All the problems of government finance are detailed as well as the notably unjust system of taxation adopted under the *ancien régime*. It was clearly Lavoisier's deep involvement in the system of taxation that led to his undoing. When he went to the guillotine in May 1794, it was not because of his scientific work, but despite it.

Minor errors creep into most scholarly books but I am surprised that Poirier did not appreciate how *recent* for Lavoisier was the recognition of the existence of a whole new state of matter, the gaseous state. The topicality for Lavoisier of gases is illustrated by the fact that, whereas in the first edition of Macquer's *Dictionary of Chemistry* (1766), only two paragraphs were devoted to 'gas', in the second edition 12 years later the same article was expanded to 100 pages. It is nonsense, therefore, to claim that in the previous century Boyle understood that "gases represented... a third important class of substances". For Boyle, and even for Hales two generations later, there was only air. This point is worth making because it was on the interpretation of the recently discovered oxygen that Lavoisier based his new interpretation of chemistry. Lavoisier was the right man in the right place at the right time.

Poirier has insisted on labelling Lavoisier anachronistically as a "biologist". Not only does he incorporate this word into the title of this revised edition but he even claims Lavoisier as "the father of modern chemistry *and biology*" (my italics). This double claim goes too far. On the other hand no one would wish to deny the importance of Lavoisier's work on respiration, which he saw as analogous to combustion. After the crude mechanical reductionism of the seventeenth century, it was an important step to interpret physiological processes in chemical terms, making possible the work of his successors, Magendie and Claude Bernard. Lavoisier also laid the foundations of organic analysis, later developed by Gay-Lussac and Liebig.

The bicentenary of Lavoisier's death in 1994 gave rise to several books and perhaps we now need to call a halt. One of the most interesting questions to emerge is to what extent the 'chemical revolution' of the eighteenth century was the work of one man. We should turn now from Lavoisier to his colleagues and assistants. He was occasionally gracious enough to acknowledge the contribution of his co-workers. A scholarly study of this research school is perhaps something we can look forward to in the next century. □

Maurice Crosland is at the Centre for History and Cultural Studies of Science, Rutherford College, University of Kent at Canterbury, Canterbury, Kent CT2 7NX, UK.

Prescriptions of the ancients

Prospecting for Drugs in Ancient and Medieval European Texts: A Scientific Approach

edited by Bart K. Holland

Harwood: 1996. Pp. 114. £39, \$65

Leslie Crombie

The investigation of indigenous medicines from all parts of the world, mostly botanical remedies, is a valuable source of 'leads' for the design of new drugs by the medicinal chemist. Knowledgeable native people who practise indigenous medicine usually know where to find the plants and how they should be used, and the medicinal chemist then has a good starting point for further development. Bart K. Holland wrote an article suggesting that prospecting for drugs among ancient and mediaeval texts might also provide a useful source of leads (*Nature* 369, 702; 1994), and this book develops the idea further.

It is at once apparent that there are more serious difficulties than arise with the investigation of modern indigenous medicine. Ancient texts cannot be questioned, unlike living practitioners. The languages used are either dead (especially Latin and Greek) or unfamiliar earlier forms of contemporary languages. Close collaboration between a classical or mediaeval scholar and a scientist is therefore essential. Deeper scholarship is needed than a mere technical ability to translate the language. The scholar needs an understanding of the medical ideas of the particular age — ideas mostly devoted to alleviating symptoms rather than curing specific conditions — and of the identification of the plants from the imprecise botanical descriptions often found. Scholars specializing in both the critical scientific and medical areas are rare, the great weight of scholarly effort being concentrated on literary aspects of ancient and mediaeval texts.



Renaissance view of nature in miniature

One of the most precious books of the European Renaissance, *The Model Book of Calligraphy*, was created at the court of the Holy Roman Emperors in the latter part of the sixteenth century. The scripts were the work of the

imperial secretary Georg Bocskay. After Bocskay's death, Emperor Rudolf II commissioned the illuminator Joris Hoefnagel to illustrate the calligrapher's work with meticulous miniatures of plants and animals.

A set of pages from the book is reproduced in *Nature Illuminated: Flora and Fauna from the Court of the Emperor Rudolf II* by Lee Hendrix and Thea Vignau-Wilberg (Thames and Hudson, £7.95).

A chapter contributed by the botanist James I. Reveal on the identification of plants in the pre-Linnean botanical literature gives an excellent account of what was known in ancient and mediaeval times, and what the difficulties can be in relating early plant descriptions to those of today. One of the great influences of classical times — and the following 16 centuries — was Pedanios Dioscorides (born about AD 40), the Cilician Greek physician to the armies of Nero, who wrote *De Materia Medica*. Dioscorides provided information on some 600 plant species along with their pharmaceutical uses. These were added to by the important Islamic writer Avicenna who wrote the *Canon of Medicine*, which was eventually translated from Arabic into Latin in the twelfth century and became a much used companion to Dioscorides' work. Such well known and translated sources form an obvious starting point in a quest for ancient medicinal remedies but there are many other less well known sources to be searched out.

So far as medical writings of the Middle Ages and Renaissance are concerned, many are available for study, if at some inconvenience to the modern reader. A very useful and scholarly account is given in a chapter by Anne Van Arsdall.

Elizabeth R. Macgill discusses a primary Renaissance source, *This Booke of Soverigne Medicines* by John Feckenham, Abbot of Westminster, which was probably used by the Benedictine Order around 1570. Macgill quotes many plant remedies for various con-

ditions and, if rather quaint, they are more or less intelligible to the modern reader, being in the English of Shakespeare.

Here is a sample: "A medicine to purge a bladder of him that cannot pisse. Take ffennell the leaves and the rootes, Allyssanders parsley, the leaves & the roote of hartstong, mayden heare, and seeth them in white wine, and give it the patient to drink, and it shall purge the bladder in a short time."

The random screening of 114,000 terrestrial plant extracts from 35,000 species by the US National Cancer Institute gave only two significant compounds — taxol and camptothecin — so the assistance of clues to narrow the choice of plants for test is welcome. Some of the really effective plant remedies of our forebears have stood the test of time and are already with us today, such as colchicine, morphine, salicylates and atropine. So the search has to be for overlooked effective botanical remedies, perhaps lost among the charlatanism, quackery and astrological beliefs of earlier days.

Such a search seems worthwhile and this book will be extremely valuable to the team best fitted for the task — the scholar with appropriate linguistic and historical qualifications working alongside the scientist, probably a pharmacist. Even with a really good lead, the long hard road to a usable modern drug is only just beginning, and may well end in failure.

Leslie Crombie is in the Department of Chemistry, University of Nottingham, Nottingham NG7 2RD, UK.

New journals

This year, *Nature's* annual new journals review supplement will appear in the issue of 11 September. Publishers and learned societies are invited to submit journals for review, as well as details of any eligible electronic journals, taking note of the following criteria:

- Journals that first appeared during or after June 1995 and issued at least four separate numbers by the end of May 1997 will be considered.
 - Journals covering any aspect of science are eligible, although those dealing with clinical medicine and pure mathematics are excluded, as are publications of abstracts.
 - Frequency of publication must be at least three times a year. The main language used must be English. Translation journals in English are, of course, eligible.
 - Deadline for submission is 6 June. When submitting journals for review, please send at least four different issues (the first, the most recent and any two others) of each title, together with full details of subscription rates.
- For further information please contact Peter Tallack, Book Reviews Editor, *Nature*, Macmillan Magazines, Porters South, Crinan Street, London N1 9XW, UK.
Tel: +44 (0)171 843 4567.
e-mail: p.tallack@nature.com

The value of the world's ecosystem services and natural capital

Robert Costanza^{*†}, Ralph d'Arge[‡], Rudolf de Groot[§], Stephen Farber^{||}, Monica Grasso[†], Bruce Hannon[‡], Karin Limburg[#], Shahid Naeem^{**}, Robert V. O'Neill^{††}, Jose Paruelo^{‡‡}, Robert G. Raskin^{§§}, Paul Sutton^{|||} & Marjan van den Belt^{¶¶}

^{*} Center for Environmental and Estuarine Studies, Zoology Department, and [†] Institute for Ecological Economics, University of Maryland, Box 38, Solomons, Maryland 20688, USA

[‡] Economics Department (emeritus), University of Wyoming, Laramie, Wyoming 82070, USA

[§] Center for Environment and Climate Studies, Wageningen Agricultural University, PO Box 9101, 6700 HB Wageningen, The Netherlands

^{||} Graduate School of Public and International Affairs, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA

[¶] Geography Department and NCSA, University of Illinois, Urbana, Illinois 61801, USA

[#] Institute of Ecosystem Studies, Millbrook, New York, USA

^{**} Department of Ecology, Evolution and Behavior, University of Minnesota, St Paul, Minnesota 55108, USA

^{††} Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA

^{‡‡} Department of Ecology, Faculty of Agronomy, University of Buenos Aires, Av. San Martin 4453, 1417 Buenos Aires, Argentina

^{§§} Jet Propulsion Laboratory, Pasadena, California 91109, USA

^{|||} National Center for Geographic Information and Analysis, Department of Geography, University of California at Santa Barbara, Santa Barbara, California 93106, USA

^{¶¶} Ecological Economics Research and Applications Inc., PO Box 1589, Solomons, Maryland 20688, USA

The services of ecological systems and the natural capital stocks that produce them are critical to the functioning of the Earth's life-support system. They contribute to human welfare, both directly and indirectly, and therefore represent part of the total economic value of the planet. We have estimated the current economic value of 17 ecosystem services for 16 biomes, based on published studies and a few original calculations. For the entire biosphere, the value (most of which is outside the market) is estimated to be in the range of US\$16–54 trillion (10¹²) per year, with an average of US\$33 trillion per year. Because of the nature of the uncertainties, this must be considered a minimum estimate. Global gross national product total is around US\$18 trillion per year.

Because ecosystem services are not fully 'captured' in commercial markets or adequately quantified in terms comparable with economic services and manufactured capital, they are often given too little weight in policy decisions. This neglect may ultimately compromise the sustainability of humans in the biosphere. The economies of the Earth would grind to a halt without the services of ecological life-support systems, so in one sense their total value to the economy is infinite. However, it can be instructive to estimate the 'incremental' or 'marginal' value of ecosystem services (the estimated rate of change of value compared with changes in ecosystem services from their current levels). There have been many studies in the past few decades aimed at estimating the value of a wide variety of ecosystem services. We have gathered together this large (but scattered) amount of information and present it here in a form useful for ecologists, economists, policy makers and the general public. From this synthesis, we have estimated values for ecosystem services per unit area by biome, and then multiplied by the total area of each biome and summed over all services and biomes.

Although we acknowledge that there are many conceptual and empirical problems inherent in producing such an estimate, we think this exercise is essential in order to: (1) make the range of potential values of the services of ecosystems more apparent; (2) establish at least a first approximation of the relative magnitude of global ecosystem services; (3) set up a framework for their further analysis; (4) point out those areas most in need of additional research; and (5) stimulate additional research and debate. Most of the problems and uncertainties we encountered indicate that our

estimate represents a minimum value, which would probably increase: (1) with additional effort in studying and valuing a broader range of ecosystem services; (2) with the incorporation of more realistic representations of ecosystem dynamics and interdependence; and (3) as ecosystem services become more stressed and 'scarce' in the future.

Ecosystem functions and ecosystem services

Ecosystem functions refer variously to the habitat, biological or system properties or processes of ecosystems. Ecosystem goods (such as food) and services (such as waste assimilation) represent the benefits human populations derive, directly or indirectly, from ecosystem functions. For simplicity, we will refer to ecosystem goods and services together as ecosystem services. A large number of functions and services can be identified^{1–4}. Reference 5 provides a recent, detailed compendium on describing, measuring and valuing ecosystem services. For the purposes of this analysis we grouped ecosystem services into 17 major categories. These groups are listed in Table 1. We included only renewable ecosystem services, excluding non-renewable fuels and minerals and the atmosphere. Note that ecosystem services and functions do not necessarily show a one-to-one correspondence. In some cases a single ecosystem service is the product of two or more ecosystem functions whereas in other cases a single ecosystem function contributes to two or more ecosystem services. It is also important to emphasize the interdependent nature of many ecosystem functions. For example, some of the net primary production in an ecosystem ends up as food, the consumption of which generates respiratory products necessary for primary production. Even though these functions and services are interdependent, in many cases they can be added because they represent 'joint products' of the ecosystem, which support human

^{*} Present address: Department of Systems Ecology, University of Stockholm, S-106 91 Stockholm, Sweden.

welfare. To the extent possible, we have attempted to distinguish joint and 'addable' products from products that would represent 'double counting' (because they represent different aspects of the same service) if they were added. It is also important to recognize that a minimum level of ecosystem 'infrastructure' is necessary in order to allow production of the range of services shown in Table 1. Several authors have stressed the importance of this 'infrastructure' of the ecosystem itself as a contributor to its total value^{6,7}. This component of the value is not included in the current analysis.

Natural capital and ecosystem services

In general, capital is considered to be a stock of materials or information that exists at a point in time. Each form of capital stock generates, either autonomously or in conjunction with services from other capital stocks, a flow of services that may be used to transform materials, or the spatial configuration of materials, to

enhance the welfare of humans. The human use of this flow of services may or may not leave the original capital stock intact. Capital stock takes different identifiable forms, most notably in physical forms including natural capital, such as trees, minerals, ecosystems, the atmosphere and so on; manufactured capital, such as machines and buildings; and the human capital of physical bodies. In addition, capital stocks can take intangible forms, especially as information such as that stored in computers and in individual human brains, as well as that stored in species and ecosystems.

Ecosystem services consist of flows of materials, energy, and information from natural capital stocks which combine with manufactured and human capital services to produce human welfare. Although it is possible to imagine generating human welfare without natural capital and ecosystem services in artificial 'space colonies', this possibility is too remote and unlikely to be of

Table 1 Ecosystem services and functions used in this study

Number	Ecosystem service*	Ecosystem functions	Examples
1	Gas regulation	Regulation of atmospheric chemical composition.	CO ₂ /O ₂ balance, O ₃ for UVB protection, and SO _x levels.
2	Climate regulation	Regulation of global temperature, precipitation, and other biologically mediated climatic processes at global or local levels.	Greenhouse gas regulation, DMS production affecting cloud formation.
3	Disturbance regulation	Capacitance, damping and integrity of ecosystem response to environmental fluctuations.	Storm protection, flood control, drought recovery and other aspects of habitat response to environmental variability mainly controlled by vegetation structure.
4	Water regulation	Regulation of hydrological flows.	Provisioning of water for agricultural (such as irrigation) or industrial (such as milling) processes or transportation.
5	Water supply	Storage and retention of water.	Provisioning of water by watersheds, reservoirs and aquifers.
6	Erosion control and sediment retention	Retention of soil within an ecosystem.	Prevention of loss of soil by wind, runoff, or other removal processes, storage of silt in lakes and wetlands.
7	Soil formation	Soil formation processes.	Weathering of rock and the accumulation of organic material.
8	Nutrient cycling	Storage, internal cycling, processing and acquisition of nutrients.	Nitrogen fixation, N, P and other elemental or nutrient cycles.
9	Waste treatment	Recovery of mobile nutrients and removal or breakdown of excess or xenic nutrients and compounds.	Waste treatment, pollution control, detoxification.
10	Pollination	Movement of floral gametes.	Provisioning of pollinators for the reproduction of plant populations.
11	Biological control	Trophic-dynamic regulations of populations.	Keystone predator control of prey species, reduction of herbivory by top predators.
12	Refugia	Habitat for resident and transient populations.	Nurseries, habitat for migratory species, regional habitats for locally harvested species, or overwintering grounds.
13	Food production	That portion of gross primary production extractable as food.	Production of fish, game, crops, nuts, fruits by hunting, gathering, subsistence farming or fishing.
14	Raw materials	That portion of gross primary production extractable as raw materials.	The production of lumber, fuel or fodder.
15	Genetic resources	Sources of unique biological materials and products.	Medicine, products for materials science, genes for resistance to plant pathogens and crop pests, ornamental species (pets and horticultural varieties of plants).
16	Recreation	Providing opportunities for recreational activities.	Eco-tourism, sport fishing, and other outdoor recreational activities.
17	Cultural	Providing opportunities for non-commercial uses.	Aesthetic, artistic, educational, spiritual, and/or scientific values of ecosystems.

* We include ecosystem 'goods' along with ecosystem services.

much current interest. In fact, one additional way to think about the value of ecosystem services is to determine what it would cost to replicate them in a technologically produced, artificial biosphere. Experience with manned space missions and with Biosphere II in Arizona indicates that this is an exceedingly complex and expensive proposition. Biosphere I (the Earth) is a very efficient, least-cost provider of human life-support services.

Thus we can consider the general class of natural capital as essential to human welfare. Zero natural capital implies zero human welfare because it is not feasible to substitute, in total, purely 'non-natural' capital for natural capital. Manufactured and human capital require natural capital for their construction⁷. Therefore, it is not very meaningful to ask the total value of natural capital to human welfare, nor to ask the value of massive, particular forms of natural capital. It is trivial to ask what is the value of the atmosphere to humankind, or what is the value of rocks and soil infrastructure as support systems. Their value is infinite in total.

However, it is meaningful to ask how changes in the quantity or quality of various types of natural capital and ecosystem services may have an impact on human welfare. Such changes include both small changes at large scales and large changes at small scales. For example, changing the gaseous composition of the global atmosphere by a small amount may have large-scale climate change effects that will affect the viability and welfare of global human populations. Large changes at small scales include, for example, dramatically changing local forest composition. These changes may dramatically alter terrestrial and aquatic ecosystems, having an impact on the benefits and costs of local human activities. In general, changes in particular forms of natural capital and ecosystem services will alter the costs or benefits of maintaining human welfare.

Valuation of ecosystem services

The issue of valuation is inseparable from the choices and decisions we have to make about ecological systems^{6,8}. Some argue that valuation of ecosystems is either impossible or unwise, that we cannot place a value on such 'intangibles' as human life, environmental aesthetics, or long-term ecological benefits. But, in fact, we do so every day. When we set construction standards for highways, bridges and the like, we value human life (acknowledged or not) because spending more money on construction would save lives. Another frequent argument is that we should protect ecosystems for purely moral or aesthetic reasons, and we do not need valuations of ecosystems for this purpose. But there are equally compelling moral arguments that may be in direct conflict with the moral argument to protect ecosystems; for example, the moral argument that no one should go hungry. Moral arguments translate the valuation and decision problem into a different set of dimensions and a different language of discourse⁶; one that, in our view, makes the problem of valuation and choice more difficult and less explicit. But moral and economic arguments are certainly not mutually exclusive. Both discussions can and should go on in parallel.

So, although ecosystem valuation is certainly difficult and fraught with uncertainties, one choice we do not have is whether or not to do it. Rather, the decisions we make as a society about ecosystems imply valuations (although not necessarily expressed in monetary terms). We can choose to make these valuations explicit or not; we can do them with an explicit acknowledgement of the huge uncertainties involved or not; but as long as we are forced to make choices, we are going through the process of valuation.

The exercise of valuing the services of natural capital 'at the margin' consists of determining the differences that relatively small changes in these services make to human welfare. Changes in quality or quantity of ecosystem services have value insofar as they either change the benefits associated with human activities or change the costs of those activities. These changes in benefits and costs either have an impact on human welfare through established markets or

through non-market activities. For example, coral reefs provide habitats for fish. One aspect of their value is to increase and concentrate fish stocks. One effect of changes in coral reef quality or quantity would be discernible in commercial fisheries markets, or in recreational fisheries. But other aspects of the value of coral reefs, such as recreational diving and biodiversity conservation, do not show up completely in markets. Forests provide timber materials through well established markets, but the associated habitat values of forests are also felt through unmarketed recreational activities. The chains of effects from ecosystem services to human welfare can range from extremely simple to exceedingly complex. Forests provide timber, but also hold soils and moisture, and create microclimates, all of which contribute to human welfare in complex, and generally non-marketed ways.

Valuation methods

Various methods have been used to estimate both the market and non-market components of the value of ecosystem services⁹⁻¹⁶. In this analysis, we synthesized previous studies based on a wide variety of methods, noting the limitations and assumptions underlying each.

Many of the valuation techniques used in the studies covered in our synthesis are based, either directly or indirectly, on attempts to estimate the 'willingness-to-pay' of individuals for ecosystem services. For example, if ecological services provided a \$50 increment to the timber productivity of a forest, then the beneficiaries of this service should be willing to pay up to \$50 for it. In addition to timber production, if the forest offered non-marketed, aesthetic, existence, and conservation values of \$70, those receiving this non-market benefit should be willing to pay up to \$70 for it. The total

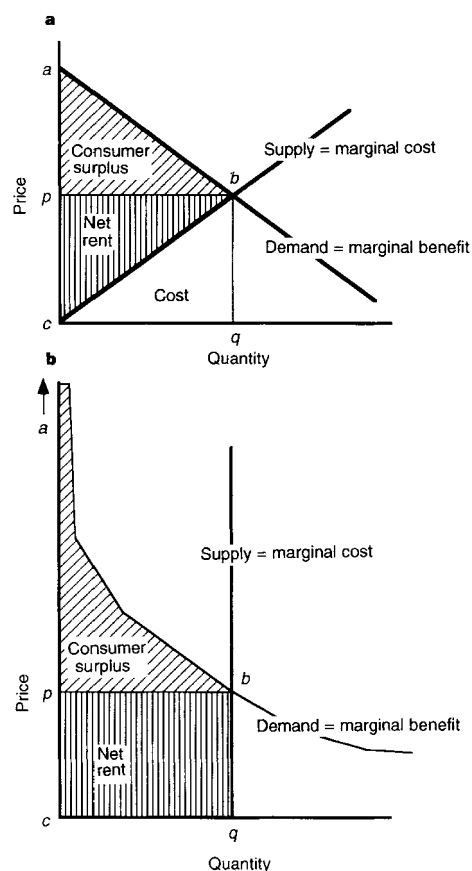


Figure 1 Supply and demand curves, showing the definitions of cost, net rent and consumer surplus for normal goods (a) and some essential ecosystem services (b). See text for further explanation.

Table 2 Summary of average global value of annual ecosystem services

Biome	Area (ha × 10 ⁶)	Ecosystem services (1994 US\$ ha ⁻¹ yr ⁻¹)																	Total value per ha (\$ ha ⁻¹ yr ⁻¹) (\$ yr ⁻¹ × 10 ⁶)	Total global flow value (\$ yr ⁻¹ × 10 ⁶)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17		
		Gas regulation	Climate regulation	Disturbance regulation	Water regulation	Water supply	Erosion control	Soil formation	Nutrient cycling	Waste treatment	Pollination	Biological control	Habitat/ refugia	Food production	Raw materials	Genetic resources	Recreation	Cultural		
Marine	36,302																		577	20,949
Open ocean	33,200	38							118			5		15	0			76	252	8,381
Coastal	3,102			88					3,677			38	8	93	4		82	62	4,062	12,568
Estuaries	180			567					21,100			78	131	521	25		381	29	22,832	4,110
Seagrass/ algae beds	200								19,002						2				19,004	3,801
Coral reefs	62			2,750				58				5	7	220	27		3,008	1	6,075	375
Shelf	2,660								1,431			39		68	2			70	1,610	4,283
Terrestrial	15,323																		804	12,319
Forest	4,855		141	2	2	3	96	10	361	87		2		43	138	16	66	2	969	4,706
Tropical	1,900		223	5	6	8	245	10	922	87				32	315	41	112	2	2,007	3,813
Temperate/boreal	2,955		88		0			10		87		4		50	25		36	2	302	894
Grass/rangelands	3,898	7	0		3		29	1		87	25	23		67		0	2		232	906
Wetlands	330	133		4,539	15	3,800				4,177			304	256	106		5/4	881	14,785	4,879
Tidal marsh/ mangroves	165			1,839						6,896			169	466	162		658		9,990	1,648
Swamps/ floodplains	165	265		7,240	30	7,600				1,659			439	47	49		491	1,761	19,560	3,231
Lakes/rivers	200				5,445	2,117				665				41			230		8,498	1,700
Desert	1,925																			
Tundra	743																			
Ice/rock	1,640																			
Cropland	1,400										14	24		54				92		128
Urban	332																			
Total	51,625	1,341	684	1,779	1,115	1,632	576	53	17075	2,277	117	417	124	1,386	721	79	815	3,015		33,288

Numbers in the body of the table are in \$ ha⁻¹ yr⁻¹. Row and column totals are the sum of the products of the per ha services in the table and the area of each biome, not the sum of the per ha services themselves. Shaded cells indicate services that do not occur or are known to be negligible. Open cells indicate lack of available information.

value of ecological services would be \$120, but the contribution to the money economy of ecological services would be \$50, the amount that actually passes through markets. In this study we have tried to estimate the total value of ecological services, regardless of whether they are currently marketed.

Figure 1 shows some of these concepts diagrammatically. Figure 1a shows conventional supply (marginal cost) and demand (marginal benefit) curves for a typical marketed good or service. The value that would show up in gross national product (GNP) is the market price p times the quantity q , or the area $pbqc$. There are three other relevant areas represented on the diagram, however. The cost of production is the area under the supply curve, cbq . The 'producer surplus' or 'net rent' for a resource is the area between the market price and the supply curve, pbc . The 'consumer surplus' or the amount of welfare the consumer receives over and above the price paid in the market is the area between the demand curve and the market price, abp . The total economic value of the resource is the sum of the producer and consumer surplus (excluding the cost of production), or the area abc on the diagram. Note that total economic value can be greater or less than the price times quantity estimates used in GNP.

Figure 1a refers to a human-made, substitutable good. Many ecosystem services are only substitutable up to a point, and their demand curves probably look more like Fig. 1b. Here the demand approaches infinity as the quantity available approaches zero (or some minimum necessary level of services), and the consumer surplus (as well as the total economic value) approaches infinity. Demand curves for ecosystem services are very difficult, if not impossible, to estimate in practice. In addition, to the extent that ecosystem services cannot be increased or decreased by actions of the economic system, their supply curves are more nearly vertical, as shown in Fig. 1b.

In this study we estimated the value per unit area of each ecosystem service for each ecosystem type. To estimate this 'unit value' we used (in order of preference) either: (1) the sum of consumer and producer surplus; or (2) the net rent (or producer surplus); or (3) price times quantity as a proxy for the economic value of the service, assuming that the demand curve for ecosystem services looks more like Fig. 1b than Fig. 1a, and that therefore the area $pbqc$ is a conservative underestimate of the area abc . We then

multiplied the unit values times the surface area of each ecosystem to arrive at global totals.

Ecosystem values, markets and GNP

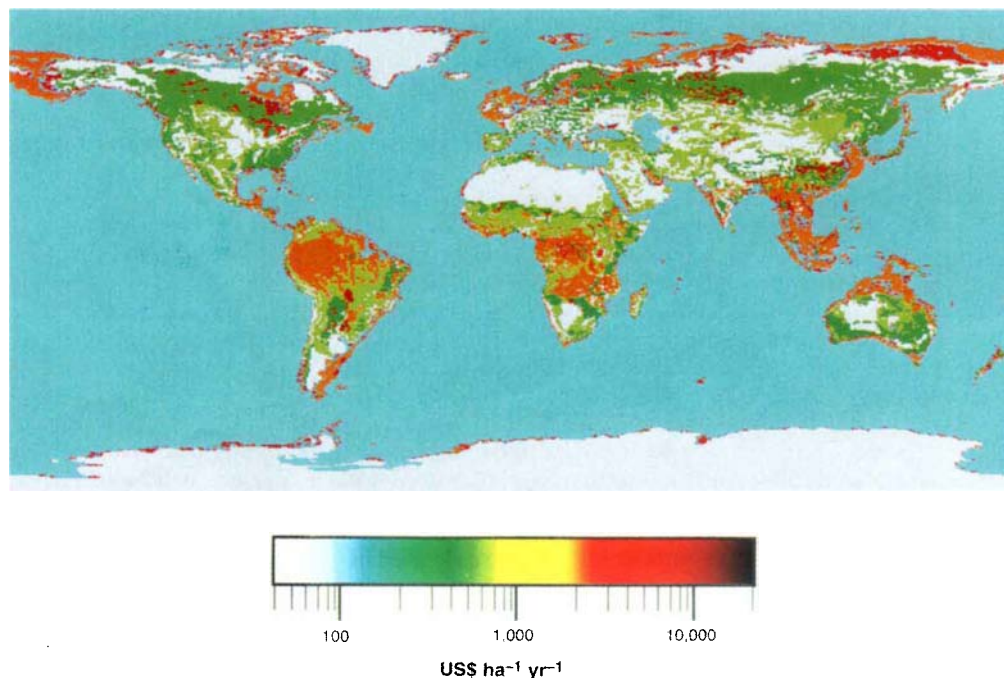
As we have noted, the value of many types of natural capital and ecosystem services may not be easily traceable through well functioning markets, or may not show up in markets at all. For example, the aesthetic enhancement of a forest may alter recreational expenditures at that site, but this change in expenditure bears no necessary relation to the value of the enhancement. Recreationists may value the improvement at \$100, but transfer only \$20 in spending from other recreational areas to the improved site. Enhanced wetlands quality may improve waste treatment, saving on potential treatment costs. For example, tertiary treatment by wetlands may save \$100 in alternative treatment. Existing treatment may cost only \$30. The treatment cost savings does not show up in any market. There is very little relation between the value of services and observable current spending behaviour in many cases.

There is also no necessary relationship between the valuation of natural capital service flows, even on the margin, and aggregate spending, or GNP, in the economy. This is true even if all capital service flows had an impact on well functioning markets. A large part of the contributions to human welfare by ecosystem services are of a purely public goods nature. They accrue directly to humans without passing through the money economy at all. In many cases people are not even aware of them. Examples include clean air and water, soil formation, climate regulation, waste treatment, aesthetic values and good health, as mentioned above.

Global land use and land cover

In order to estimate the total value of ecosystem services, we needed estimates of the total global extent of the ecosystems themselves. We devised an aggregated classification scheme with 16 primary categories as shown in Table 2 to represent current global land use. The major division is between marine and terrestrial systems. Marine was further subdivided into open ocean and coastal, which itself includes estuaries, seagrass/algae beds, coral reefs, and shelf systems. Terrestrial systems were broken down into two types of forest (tropical and temperate/boreal), grasslands/rangelands, wetlands, lakes/streams, desert, tundra, ice/rock, cropland, and urban. Primary

Figure 2 Global map of the value of ecosystem services. See Supplementary Information and Table 2 for details.



data were from ref. 17 as summarized in ref. 4 with additional information from a number of sources^{18–22}. We also used data from ref. 23, as a cross-check on the terrestrial estimates and refs 24 and 25 as a check on the marine estimates. The 32 landcover types of ref. 17 were recategorized for Table 2 and Fig. 2. The major assumptions were: (1) chaparral and steppe were considered rangeland and combined with grasslands; and (2) a variety of tropical forest and woodland types were combined into 'tropical forests'.

Synthesis

We conducted a thorough literature review and synthesized the information, along with a few original calculations, during a one-week intensive workshop at the new National Center for Ecological Analysis and Synthesis (NCEAS) at the University of California at Santa Barbara. Supplementary Information lists the primary results for each ecosystem service and biome. Supplementary Information includes all the estimates we could identify from the literature (from over 100 studies), their valuation methods, location and stated value. We converted each estimate into 1994 US\$ ha⁻¹ yr⁻¹ using the USA consumer price index and other conversion factors as needed. These are listed in the notes to the Supplementary Information. For some estimates we also converted the service estimate into US\$ equivalents using the ratio of purchasing power GNP per capita for the country of origin to that of the USA. This was intended to adjust for income effects. Where possible the estimates are stated as a range, based on the high and low values found in the literature, and an average value, with annotated comments as to methods and assumptions. We also included in the Supplementary Information some estimates from the literature on 'total ecosystem value', mainly using energy analysis techniques¹⁰. We did not include these estimates in any of the totals or averages given below, but only for comparison with the totals from the other techniques. Interestingly, these different methods showed fairly close agreement in the final results.

Each biome and each ecosystem service had its special considerations. Detailed notes explaining each biome and each entry in Supplementary Information are given in notes following the table. More detailed descriptions of some of the ecosystems, their services, and general valuation issues can be found in ref. 5. Below we briefly discuss some general considerations that apply across the board.

Sources of error, limitations and caveats

Our attempt to estimate the total current economic value of ecosystem services is limited for a number of reasons, including:

- (1) Although we have attempted to include as much as possible, our estimate leaves out many categories of services, which have not yet been adequately studied for many ecosystems. In addition, we could identify no valuation studies for some major biomes (desert, tundra, ice/rock, and cropland). As more and better information becomes available we expect the total estimated value to increase.
- (2) Current prices, which form the basis (either directly or indirectly) of many of the valuation estimates, are distorted for a number of reasons, including the fact that they exclude the value of ecosystem services, household labour and the informal economy. In addition to this, there are differences between total value, consumer surplus, net rent (or producer surplus) and $p \times q$, all of which are used to estimate unit values (see Fig. 1).
- (3) In many cases the values are based on the current willingness-to-pay of individuals for ecosystem services, even though these individuals may be ill-informed and their preferences may not adequately incorporate social fairness, ecological sustainability and other important goals¹⁶. In other words, if we actually lived in a world that was ecologically sustainable, socially fair and where everyone had perfect knowledge of their connection to ecosystem services, both market prices and surveys of willingness-to-pay would yield very different results than they currently do, and the value of ecosystem services would probably increase.

(4) In calculating the current value, we generally assumed that the demand and supply curves look something like Fig. 1a. In reality, supply curves for many ecosystem services are more nearly inelastic vertical lines, and the demand curves probably look more like Fig. 1b, approaching infinity as quantity goes to zero. Thus the consumer and producer surplus and thereby the total value of ecosystem services would also approach infinity.

(5) The valuation approach taken here assumes that there are no sharp thresholds, discontinuities or irreversibilities in the ecosystem response functions. This is almost certainly not the case. Therefore this valuation yields an underestimate of the total value.

(6) Extrapolation from point estimates to global totals introduces error. In general, we estimated unit area values for the ecosystem services (in \$ ha⁻¹ yr⁻¹) and then multiplied by the total area of each biome. This can only be considered a crude first approximation and can introduce errors depending on the type of ecosystem service and its spatial heterogeneity.

(7) To avoid double counting, a general equilibrium framework that could directly incorporate the interdependence between ecosystem functions and services would be preferred to the partial equilibrium framework used in this study (see below).

(8) Values for individual ecosystem functions should be based on sustainable use levels, taking account of both the carrying capacity for individual functions (such as food-production or waste recycling) and the combined effect of simultaneous use of more functions. Ecosystems should be able to provide all the functions listed in Table 1 simultaneously and indefinitely. This is certainly not the case for some current ecosystem services because of overuse at existing prices.

(9) We have not incorporated the 'infrastructure' value of ecosystems, as noted above, leading to an underestimation of the total value.

(10) Inter-country comparisons of valuation are affected by income differences. We attempted to address this in some cases using the relative purchasing power GNP per capita of the country relative to the USA, but this is a very crude way to make the correction.

(11) In general, we have used annual flow values and have avoided many of the difficult issues involved with discounting future flow values to arrive at a net present value of the capital stock. But a few estimates in the literature were stated as stock values, and it was necessary to assume a discount rate (we used 5%) in order to convert them into annual flows.

(12) Our estimate is based on a static 'snapshot' of what is, in fact, a complex, dynamic system. We have assumed a static and 'partial equilibrium' model in the sense that the value of each service is derived independently and added. This ignores the complex interdependencies between the services. The estimate could also change drastically as the system moved through critical non-linearities or thresholds. Although it is possible to build 'general equilibrium' models in which the value of all ecosystem services are derived simultaneously with all other values, and to build dynamic models that can incorporate non-linearities and thresholds, these models have rarely been attempted at the scale we are discussing. They represent the next logical step in deriving better estimates of the value of ecosystem services.

We have tried to expose these various sources of uncertainty wherever possible in Supplementary Information and its supporting notes, and state the range of relevant values. In spite of the limitations noted above, we believe it is very useful to synthesize existing valuation estimates, if only to determine a crude, initial magnitude. In general, because of the nature of the limitations noted, we expect our current estimate to represent a minimum value for ecosystem services.

Total global value of ecosystem services

Table 2 is a summary of the results of our synthesis. It lists each of the major biomes along with their current estimated global surface

area, the average (on a per hectare basis) of the estimated values of the 17 ecosystem services we have identified from Supplementary Information, and the total value of ecosystem services by biome, by service type and for the entire biosphere.

We estimated that at the current margin, ecosystems provide at least US\$33 trillion dollars worth of services annually. The majority of the value of services we could identify is currently outside the market system, in services such as gas regulation (US\$1.3 trillion yr⁻¹), disturbance regulation (US\$1.8 trillion yr⁻¹), waste treatment (US\$2.3 trillion yr⁻¹) and nutrient cycling (US\$17 trillion yr⁻¹). About 63% of the estimated value is contributed by marine systems (US\$20.9 trillion yr⁻¹). Most of this comes from coastal systems (US\$10.6 trillion yr⁻¹). About 38% of the estimated value comes from terrestrial systems, mainly from forests (US\$4.7 trillion yr⁻¹) and wetlands (US\$4.9 trillion yr⁻¹).

We estimated a range of values whenever possible for each entry in Supplementary Information. Table 2 reports only the average values. Had we used the low end of the range in Supplementary Information, the global total would have been around US\$19 trillion. If we eliminate nutrient cycling, which is the largest single service, estimated at US\$17 trillion, the total annual value would be around US\$16 trillion. Had we used the high end for all estimates, along with estimating the value of desert, tundra and ice/rock as the average value of rangelands, the estimate would be around US\$54 trillion. So the total range of annual values we estimated were from US\$16–\$54 trillion. This is not a huge range, but other sources of uncertainty listed above are much more critical. It is important to emphasize, however, that despite the many uncertainties included in this estimate, it is almost certainly an underestimate for several reasons, as listed above.

There have been very few previous attempts to estimate the total global value of ecosystem services with which to compare these results. We identified two, based on completely different methods and assumptions, both from each other and from the methods used in this study. They thus provide an interesting check.

One was an early attempt at a static general equilibrium input–output model of the globe, including both ecological and economic processes and commodities^{26,27}. This model divided the globe into 9 commodities or product groups and 9 processes, two of which were 'economic' (urban and agriculture) and 7 of which were 'ecological', including both terrestrial and marine systems. Data were from about 1970. Although this was a very aggregated breakdown and the data was of only moderate quality, the model produced a set of 'shadow prices' and 'shadow values' for all the flows between processes, as well as the net outputs from the system, which could be used to derive an estimate of the total value of ecosystem services. The input–output format is far superior to the partial equilibrium format we used in this study for differentiating gross from net flows and avoiding double counting. The results yielded a total value of the net output of the 7 global ecosystem processes equal to the equivalent of US\$9.4 trillion in 1972. Converted to 1994 US\$ this is about \$34 trillion, surprisingly close to our current average estimate. This estimate broke down into US\$11.9 trillion (or 35%) from terrestrial ecosystem processes and US\$22.1 trillion (or 65%) from marine processes, also very close to our current estimate. World GNP in 1970 was about \$14.3 trillion (in 1994 US\$), indicating a ratio of total ecosystem services to GNP of about 2.4 to 1. The current estimate has a corresponding ratio of 1.8 to 1.

A more recent study²⁸ estimated a 'maximum sustainable surplus' value of ecosystem services by considering ecosystem services as one input to an aggregate global production function along with labour and manufactured capital. Their estimates ranged from US\$3.4 to US\$17.6 trillion yr⁻¹, depending on various assumptions. This approach assumed that the total value of ecosystem services is limited to that which has an impact on marketed value, either directly or indirectly, and thus cannot exceed the total world GNP of about US\$18 trillion. But, as we have pointed out, only a fraction of

ecosystem services affects private goods traded in existing markets, which would be included in measures such as GNP. This is a subset of the services we estimated, so we would expect this estimate to undervalue total ecosystem services.

The results of both of these studies indicate, however, that our current estimate is at least in approximately the same range. As we have noted, there are many limitations to both the current and these two previous studies. They are all only static snapshots of a biosphere that is a complex, dynamic system. The obvious next steps include building regional and global models of the linked ecological economic system aimed at a better understanding of both the complex dynamics of physical/biological processes and the value of these processes to human well-being^{29,30}. But we do not have to wait for the results of these models to draw the following conclusions.

Discussion

What this study makes abundantly clear is that ecosystem services provide an important portion of the total contribution to human welfare on this planet. We must begin to give the natural capital stock that produces these services adequate weight in the decision-making process, otherwise current and continued future human welfare may drastically suffer. We estimate in this study that the annual value of these services is US\$16–54 trillion, with an estimated average of US\$33 trillion. The real value is almost certainly much larger, even at the current margin. US\$33 trillion is 1.8 times the current global GNP. One way to look at this comparison is that if one were to try to replace the services of ecosystems at the current margin, one would need to increase global GNP by at least US\$33 trillion, partly to cover services already captured in existing GNP and partly to cover services that are not currently captured in GNP. This impossible task would lead to no increase in welfare because we would only be replacing existing services, and it ignores the fact that many ecosystem services are literally irreplaceable.

If ecosystem services were actually paid for, in terms of their value contribution to the global economy, the global price system would be very different from what it is today. The price of commodities using ecosystem services directly or indirectly would be much greater. The structure of factor payments, including wages, interest rates and profits would change dramatically. World GNP would be very different in both magnitude and composition if it adequately incorporated the value of ecosystem services. One practical use of the estimates we have developed is to help modify systems of national accounting to better reflect the value of ecosystem services and natural capital. Initial attempts to do this paint a very different picture of our current level of economic welfare than conventional GNP, some indicating a levelling of welfare since about 1970 while GNP has continued to increase^{31–33}. A second important use of these estimates is for project appraisal, where ecosystem services lost must be weighed against the benefits of a specific project⁸. Because ecosystem services are largely outside the market and uncertain, they are too often ignored or undervalued, leading to the error of constructing projects whose social costs far outweigh their benefits.

As natural capital and ecosystem services become more stressed and more 'scarce' in the future, we can only expect their value to increase. If significant, irreversible thresholds are passed for irreplaceable ecosystem services, their value may quickly jump to infinity. Given the huge uncertainties involved, we may never have a very precise estimate of the value of ecosystem services. Nevertheless, even the crude initial estimate we have been able to assemble is a useful starting point (we stress again that it is only a starting point). It demonstrates the need for much additional research and it also indicates the specific areas that are most in need of additional study. It also highlights the relative importance of ecosystem services and the potential impact on our welfare of continuing to squander them. □

1. de Groot, R. S. Environmental functions as a unifying concept for ecology and economics. *Environmentalist* 7, 105–109 (1987).
2. Turner, R. K. *Economics, Growth and Sustainable Environments* (eds Collard, D. et al.) (Macmillan, London, 1988).
3. Turner, R. K. Economics of wetland management. *Ambio* 20, 59–63 (1991).
4. de Groot, R. S. *Functions of Nature: Evaluation of Nature in Environmental Planning, Management, and Decision Making* (Wolters-Noordhoff, Groningen, 1992).
5. Daily, G. (ed.) *Nature's Services: Societal Dependence on Natural Ecosystems* (Island, Washington DC, 1997).
6. Turner, R. K. & Pearce, D. in *Economics and Ecology: New Frontiers and Sustainable Development* (ed. Barbier, E. D.) 177–194 (Chapman and Hall, London, 1993).
7. Costanza, R. & Daly, H. E. Natural capital and sustainable development. *Conserv. Biol.* 6, 37–46 (1992).
8. Bingham, G. et al. Issues in ecosystem valuation: improving information for decision making. *Ecol. Econ.* 14, 73–90 (1995).
9. Mitchell, R. C. & Carson, R. T. *Using Surveys to Value Public Goods: the Contingent Valuation Method* (Resources for the Future, Washington DC, 1989).
10. Costanza, R., Farber, S. C. & Maxwell, J. Valuation and management of wetlands ecosystems. *Ecol. Econ.* 1, 335–361 (1989).
11. Dixon, J. A. & Sherman, P. B. *Economics of Protected Areas* (Island, Washington DC, 1990).
12. Barde, J.-P. & Pearce, D. W. *Valuing the Environment: Six Case Studies* (Earthscan, London, 1991).
13. Aylward, B. A. & Barbier, E. B. Valuing environmental functions in developing countries. *Biodiv. Cons.* 1, 34 (1992).
14. Pearce, D. *Economic Values and the Natural World* (Earthscan, London, 1993).
15. Goulder, L. H. & Kennedy, D. in *Nature's Services: Societal Dependence on Natural Ecosystems* (ed. Daily, G.) 23–48 (Island, Washington DC, 1997).
16. Costanza, R. & Folke, C. in *Nature's Services: Societal Dependence on Natural Ecosystems* (ed. Daily, G.) 49–70 (Island, Washington DC, 1997).
17. Matthews, E. Global vegetation and land-use: new high-resolution data bases for climate studies. *J. Clim. Appl. Meteorol.* 22, 474–487 (1983).
18. Deevey, E. S. Mineral cycles. *Sci. Am.* 223, 148–158 (1970).
19. Ehrlich, R., Ehrlich, A. H. & Holdren, J. P. *Ecoscience: Population, Resources, Environment* (W.H. Freeman, San Francisco, 1977).
20. Ryther, J. H. Photosynthesis and fish production in the sea. *Science* 166, 72–76 (1969).
21. United Nations Environmental Programme *First Assessment Report, Intergovernmental Panel on*

- Climate Change* (United Nations, New York, 1990).
22. Whittaker, R. H. & Likens, G. E. in *Primary Production of the Biosphere* (eds Lieth, H. & Whittaker, R. H.) 305–328 (Springer, New York, 1975).
23. Bailey, R. G. *Ecosystem Geography* (Springer, New York, 1996).
24. Houde, E. D. & Rutherford, E. S. Recent trends in estuarine fisheries: predictions of fish production and yield. *Estuaries* 16, 161–176 (1993).
25. Pauly, D. & Christensen, V. Primary production required to sustain global fisheries. *Nature* 374, 255–257 (1995).
26. Costanza, R. & Neil, C. in *Energy and Ecological Modeling* (eds Mitsch, W. J., Bosserman, R. W. & Klopatek, J. M.) 745–755 (Elsevier, New York, 1981).
27. Costanza, R. & Hannon, B. M. in *Network Analysis of Marine Ecosystems: Methods and Applications* (eds Wulff, F., Field, J. G. & Mann, K. H.) 90–115 (Springer, Heidelberg, 1989).
28. Alexander, A., List, J., Margolis, M. & d'Arge, R. Alternative methods of valuing global ecosystem services. *Ecol. Econ.* (submitted).
29. Costanza, R., Wainger, L., Folke, C. & Mäler, K.-G. Modeling complex ecological economic systems: toward an evolutionary, dynamic understanding of people and nature. *BioScience* 43, 545–555 (1993).
30. Bockstael, N. et al. Ecological economic modeling and valuation of ecosystems. *Ecol. Econ.* 14, 143–159 (1995).
31. Daly, H. E. & Cobb, J. *For the Common Good: Redirecting the Economy Towards Community, the Environment, and a Sustainable Future* (Beacon, Boston, 1989).
32. Cobb, C. & Cobb, J. *The Green National Product: a Proposed Index of Sustainable Economic Welfare* (Univ. Press of America, New York, 1994).
33. Max-Neef, M. Economic growth and quality of life: a threshold hypothesis. *Ecol. Econ.* 15, 115–118 (1995).

Acknowledgements. S. Carpenter was instrumental in encouraging the project. M. Grasso did the initial identification and collection of literature sources. We thank S. Carpenter, G. Daily, H. Daly, A. M. Freeman, N. Myers, C. Perrings, D. Pimentel, S. Pimm and S. Postel for helpful comments on earlier drafts. This project was sponsored by the National Center for Ecological Analysis and Synthesis (NCEAS), an NSF-funded Center at the University of California at Santa Barbara. The authors met during the week of June 17–21, 1996 to do the major parts of the synthesis activities. The idea for the study emerged at a meeting of the Pew Scholars in New Hampshire in October 1995.

Correspondence and requests for materials should be addressed to R.C. (e-mail: costza@cbl.cees.edu).

Supplementary Information is on www.nature.com. Paper copies are available from Mary Sheehan at the London editorial office of *Nature*.

KNOW YOUR COPY RIGHTS RESPECT OURS

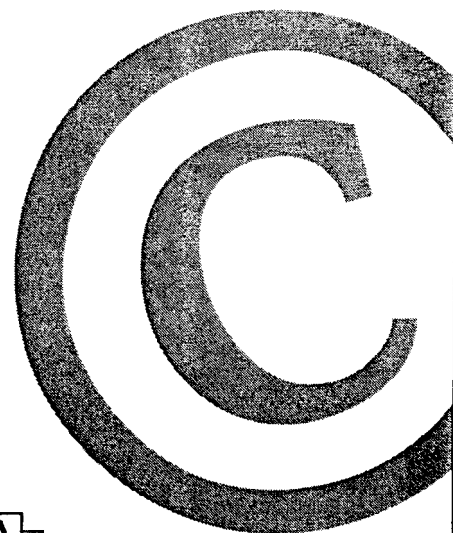
The publication you are reading is protected by copyright law. This means that the publisher could take you and your employer to court and claim heavy legal damages if you make unauthorised photocopies from these pages.

Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

The Copyright Licensing Agency (CLA) is an organisation which issues licences to bring photocopying within the law. It has designed licensing services to cover all kinds of special needs in business, education, and government.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP

Telephone: 0171 436 5931
Fax: 0171 436 3986

Absence of a magnetic-field signature in plasma-wave observations at Callisto

D. A. Gurnett*, W. S. Kurth*, A. Roux† & S. J. Bolton‡

* Department of Physics and Astronomy, University of Iowa, Iowa City, Iowa 52242, USA

† Centre d'Etudes des Environnements Terrestre et Planétaires, Université Versailles Saint Quentin, 10/12 Avenue de l'Europe, 78140 Velizy, France

‡ Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, California 91109, USA

The galilean moons of Jupiter are substantial bodies—three of them are larger than the Earth's Moon, and one is larger than Mercury—yet little has been known about them until very recently. The discovery of a magnetosphere¹ and magnetic field² associated with Ganymede was surprising, and raised the possibility that other galilean satellites, particularly Callisto (which is the second largest after Ganymede), also might have an internally generated magnetic field. Here we report observations of plasma waves around Callisto, detected during the recent fly-by of the Galileo spacecraft. The nature of the plasma waves indicates that Callisto, unlike Ganymede, does not have a magnetosphere or an internal magnetic field. The electron density near Callisto, however, is substantially higher than that in Jupiter's magnetosphere at this orbital radius, indicating that Callisto is a significant source of locally generated plasma. This plasma most probably comes from a tenuous atmosphere around Callisto, which may be similar to the hydrogen cloud around Ganymede, as the electron densities are somewhat comparable.

The trajectory for the 4 November Callisto fly-by is shown in Fig. 1. This first close fly-by of Callisto was designed to provide measurements in the wake created by the co-rotating plasma of Jupiter³, which flows by Callisto at a speed of roughly 200 km s^{-1} (ref. 4). The closest approach occurred at 13:34:30 UT (universal time) at an altitude of 1,129 km and a radial distance of $1.47 R_C$ (where Callisto's radius, R_C , is 2,403 km). As can be seen, the trajectory passed directly through the geometric wake of Callisto. The time required to cross the wake is about 10 minutes.

Spectrograms of the electric and magnetic field intensities obtained from the plasma-wave instrument for a 2.5-hour interval around the time of closest approach are shown in Fig. 2. For a description of the plasma-wave instrument, see ref. 5. As can be seen in the bottom panel of Fig. 2, there are no magnetic wave fields associated with Callisto. The absence of a magnetic response is in sharp contrast to the Ganymede fly-by¹, where strong electromagnetic emissions propagating in a mode known as the whistler mode⁶ were observed at frequencies up to about 5 kHz. The absence of whistler-mode emissions at Callisto can be explained in part by magnetometer measurements⁷, which show that the magnetic field near Callisto is very weak, $B \leq 35 \text{ nT}$. As the whistler mode cannot propagate at frequencies above the electron cyclotron frequency, $f_c = 28 \text{ MHz}$, where B is the magnetic field strength in nanoteslas, whistler-mode emissions would not be expected in the vicinity of Callisto at frequencies above about 1 kHz. However, the weak magnetic field does not explain the absence of whistler-mode emissions at frequencies below 1 kHz. At Ganymede, strong whistler-mode emissions were observed down to frequencies as low as a few hundred hertz. The absence of similar low-frequency whistler-mode emissions at Callisto indicates that Callisto's interaction with the jovian magnetosphere is fundamentally different from at Ganymede. Most probably this difference is caused by the absence of an internally generated magnetic field.

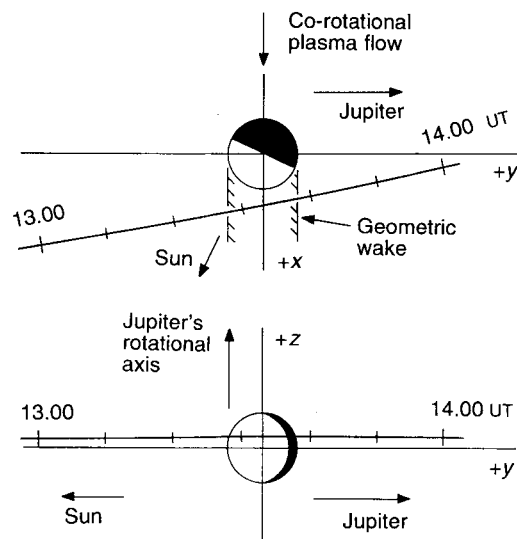


Figure 1 The Galileo trajectory during the 4 November 1996 Callisto fly-by. The Callisto-centred coordinate system has the +z-axis aligned parallel to Jupiter's rotational axis and the +x-axis parallel to the nominal co-rotational plasma flow induced by Jupiter's rotation. The +y-axis completes the right-hand coordinate system.

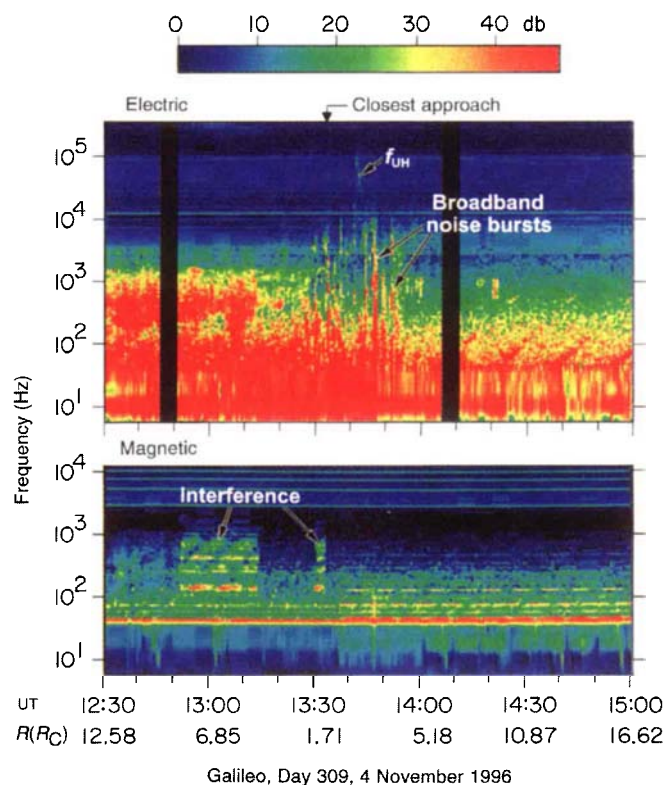


Figure 2 Frequency-time spectrograms of the electric (top panel) and magnetic (bottom panel) field intensities detected by the Galileo plasma wave instrument. The radial distance, R , from the centre of Callisto is given at the bottom of the plot in Callisto radii, R_C . The time of closest approach is indicated by the arrow at the top. Various horizontal lines, particularly in the magnetic field spectrogram, are caused by spacecraft-generated interference. The broadband magnetic field enhancements below about 10^3 Hz from about 12:52 to 13:12 UT, and again from about 13:29 to 13:33 UT, are caused by interference from the Ultraviolet Spectrometer instrument.

Inspection of the electric field spectrogram in Fig. 2 shows a series of noise bursts, from about 13:28 to 13:52 UT, which clearly stand out above the prevailing noise background. These noise bursts extend over a frequency range from about 10^2 to 10^4 Hz and seem to be associated with Callisto. No comparable noise bursts have been observed during other crossings of Callisto's orbit. Comparisons with the magnetic field measurements of ref. 7 clearly show that the bursts of electric field noise are closely correlated with magnetic field rotations. For example, the noise bursts at 13:33, 13:47 and 13:52 UT all correspond to abrupt changes in the B_x and B_y magnetic field components, changes that are indicative of field-aligned currents. Khurana *et al.*⁷ have suggested that some of these field rotations (particularly from 13:40 to 13:52 UT) may be related to field-aligned currents flowing along the boundary of Jupiter's plasma sheet. Galileo was just crossing into the plasma sheet at the time of the Callisto fly-by. Data from the forthcoming Galileo fly-bys of Callisto should allow us to identify the origin of these field-aligned currents.

In addition to the bursts of electric field noise, a weak narrow-band emission can be seen in the electric field spectrogram drifting downwards in frequency from about 90 kHz at 13:41:30 UT to about 20 kHz at 13:43:30 UT. This emission intensifies at about 13:42:00 UT, causing a well defined enhancement in the spectrum at about 50 kHz. From similar observations at Jupiter and other planets^{8–11}, we can identify this emission as an electrostatic oscillation that occurs at the upper hybrid resonance frequency. For a discussion of this resonance, see ref. 6. Upper hybrid emissions provide a very accurate measurement of the local electron density. The upper hybrid resonance frequency is given by $f_{UH} = (f_p^2 + f_c^2)^{1/2}$, where $f_p = 8,980 \sqrt{N}$ Hz is the electron plasma frequency⁶ and N is the electron density in cm^{-3} . Solving the above equation for the electron density gives $N = (f_{UH}^2 - f_c^2)/(8,980)^2 \text{ cm}^{-3}$. For $f_c = 1 \text{ kHz}$ (that is, $B \approx 35 \text{ nT}$), the corresponding electron densities vary from about 100 cm^{-3} ($f_{UH} = 90 \text{ kHz}$) at 13:41:30 UT to about 5 cm^{-3} ($f_{UH} = 20 \text{ kHz}$) at 13:43:30 UT. As $f_c \ll f_{UH}$, the magnetic field plays only a minor role in the calculation of electron density. The electron density associated with the intensification at 13:42:00 UT ($f_{UH} = 50 \text{ kHz}$) is 31 cm^{-3} .

Two factors lead us to believe that these plasma densities are associated with Callisto. First, even though upper hybrid emissions are a common features of the jovian magnetosphere, these emissions usually occur in the inner regions of the magnetosphere very close ($\pm 2^\circ$) to the magnetic equator¹². The Callisto fly-by occurred at a magnetic latitude of -7° , well outside the region where upper hybrid emissions are usually observed. Second, the electron densities, $5\text{--}100 \text{ cm}^{-3}$, are much higher than would be expected in the jovian magnetosphere at Callisto's orbit. For example, from electron density measurements made by Voyagers 1 and 2 (refs 4, 11), the maximum electron density at Callisto's orbit is expected to be less than 1 cm^{-3} . Thus, the plasma detected near Callisto almost certainly originates from Callisto. Because a plasma is usually produced by ionizing a neutral gas, such large plasma densities, up to 100 cm^{-3} , imply that a significant neutral atmosphere must exist around Callisto, possibly similar to the hydrogen exosphere recently discovered around Ganymede¹³.

From the above description, it is clear that the plasma-wave signature of Callisto is quite different from that of Ganymede. There is no evidence of the types of plasma waves that are typically associated with a magnetized planet. Instead, the waves observed are more like those found near an unmagnetized object such as the Earth's Moon¹⁴ or a comet¹⁵. Our results support the measurements of refs 7 and 16, which show that Callisto has little, if any, internally generated magnetic field. $\square$

Received 2 January; accepted 18 March 1997.

1. Gurnett, D. A., Kurth, W. S., Roux, A., Bolton, S. J. & Kennel, C. F. Evidence for a magnetosphere at Ganymede from plasma-wave observations by the Galileo spacecraft. *Nature* **384**, 535–537 (1996).
2. Kivelson, M. G. *et al.* Discovery of Ganymede's magnetic field by the Galileo spacecraft. *Nature* **384**, 537–541 (1996).
3. Johnson, T. V., Yeates, C. M. & Young, R. The Galileo mission overview. *Space Sci. Rev.* **60**, 3–21 (1992).

4. McNutt, R. L., Belcher, J. W. Jr & Bridge, H. S. Positive ion observations in the middle magnetosphere of Jupiter. *J. Geophys. Res.* **86**, 8319–8342 (1981).
5. Gurnett, D. A. *et al.* The Galileo plasma wave investigation. *Space Sci. Rev.* **60**, 341–355 (1992).
6. Stix, T. H. *The Theory of Plasma Waves* 12 (McGraw-Hill, New York, 1962).
7. Khurana, K. K., Kivelson, M. G., Russell, C. T., Walker, R. J. & Southwood, D. J. Absence of an internal magnetic field at Callisto. *Nature* **387**, 262–264 (1997).
8. Walsh, D., Haddock, T. F. & Schulte, H. F. Cosmic radio intensities at 1.225 and 2.0 MC measured up to an altitude of 1700 km. *Space Res.* **4**, 935–959 (1964).
9. Mosier, S. R., Kaiser, M. L. & Brown, L. W. Observations of noise bands associated with the upper hybrid resonance by the IMP 6 radio astronomy experiment. *J. Geophys. Res.* **78**, 1673–1677 (1973).
10. Warwick, J. W. *et al.* Voyager 1 planetary radio astronomy observations near Jupiter. *Science* **204**, 995–998 (1979).
11. Gurnett, D. A., Scarf, F. L., Kurth, W. S., Shaw, R. R. & Poynter, R. L. Determination of Jupiter's electron density profile from plasma wave observations. *J. Geophys. Res.* **86**, 8199–8212 (1981).
12. Kurth, W. S., Barbosa, D. D., Gurnett, D. A. & Scarf, F. L. Electrostatic waves in the jovian magnetosphere. *Geophys. Res. Lett.* **7**, 57–60 (1980).
13. Barth, C. *et al.* Galileo Ultraviolet Spectrometer observations of atomic hydrogen in the atmosphere of Ganymede (abstr.). *Eos* **77**, F430 (1996).
14. Kellogg, P. J. *et al.* Observations of plasma waves during a traversal of the Moon's wake. *J. Geophys. Res.* **23**, 1267–1270 (1996).
15. Scarf, F. L. *et al.* Plasma wave observations at comet Giacobini–Zinner. *Science* **232**, 377–381 (1986).
16. Anderson, J. D., Lau, E. L., Sjogren, W. L., Schubert, G. & Moore, W. B. Gravitational evidence for an undifferentiated Callisto. *Nature* **387**, 264–266 (1997).

Acknowledgements. We thank the Galileo project team at the Jet Propulsion Laboratory for their considerable efforts in obtaining these data. We also thank K. Khurana for providing the magnetic field data before publication, and L. Granroth and J. Groene for their assistance in the data processing at the University of Iowa. This research was supported by NASA through the Jet Propulsion Laboratory, and by the Centre National d'Etudes Spatiales (France).

Correspondence should be addressed to D.A.G. (e-mail: donald-gurnett@uiowa.edu).

Absence of an internal magnetic field at Callisto

K. K. Khurana*, M. G. Kivelson*†, C. T. Russell*†, R. J. Walker* & D. J. Southwood‡

* Institute of Geophysics and Planetary Physics, † Department of Earth and Space Sciences, University of California, Los Angeles, California 90095, USA

‡ Department of Physics, Imperial College of Science, Technology, and Medicine, London SW7 2BZ, UK

Little is known about the internal properties of Callisto—the outermost of Jupiter's four large galilean moons—other than the average density (about 1.8 g cm^{-3}). The recent unexpected discovery^{1–4} that Ganymede, and perhaps Io, has an internally generated magnetic field, combined with gravity results^{5,6} suggesting that both Ganymede and Io are internally differentiated with metallic cores and rocky mantles, has heightened anticipation of the results obtained by the Galileo spacecraft in its recent fly-by of Callisto. Here we report that the spacecraft, passing the moon at a distance of only $\sim 1,100 \text{ km}$ from the surface, detected only a small enhancement of the field strength ($\sim 7 \text{ nT}$), which may be related to changes in the jovian plasma environment caused by Callisto⁷. Callisto does not have an internally generated magnetic field.

On 4 November 1996, Galileo passed close to Callisto. Low-resolution magnetic-field⁸ ($\Delta t = 24 \text{ s}$) and particle data were acquired continuously around the time of closest approach. In addition, 45 minutes of full-resolution magnetic-field ($\Delta t = 0.33 \text{ s}$) and plasma data were tape-recorded on board the spacecraft in the close vicinity of Callisto. These data have now been returned to Earth. Here we summarize the magnetometer observations from this close pass.

The closest approach occurred near 13:34:30 UT at a distance of 1.47 times the radius of Callisto ($1.47 R_C$; $R_C = 2,403 \text{ km}$) at a latitude of 13° with respect to Callisto's equator. The jovian plasma at the radial distance of Callisto moves⁹ slower than the co-rotational velocity of 330 km s^{-1} but has a velocity in excess of 200 km s^{-1} ; it therefore continually overtakes Callisto whose orbital speed is 8.2 km s^{-1} . Thus, a wake is formed ahead of Callisto in the downstream plasma. Galileo passed through this wake region at a distance of $\sim 1,100 \text{ km}$ from the surface of Callisto.

A five-hour (half jovian day) plot of the merged high- and low-

resolution magnetic field data in System III coordinates¹⁰ is shown in Fig. 1. The variations of the vector magnetic field along the trajectory were caused principally by the motion of Jupiter's current sheet controlled by the rotating tilted jovian dipole. When the spacecraft was located above (below) the current sheet, the radial component of the field (B_r ; see Fig. 1 legend) was positive (negative). During 12:00–13:00 UT and 16:00–17:00 UT, B_r remained essentially constant in magnitude in the regions outside the plasma sheet referred to as lobes. As the spacecraft travelled through the current sheet, B_r varied rapidly from 14:00 to 16:00 UT. The Callisto encounter occurred at the time when the spacecraft was beginning to descend into the current sheet. Using the thin current sheet approximation introduced by Vasyliunas¹¹, the ratio of the thermal pressure to the magnetic pressure (β) at any location in the magnetodisk can be calculated from $\beta = (B_L^2 - B^2)/B^2$, where B_L is the field strength in the lobe and B is the field strength at that location. Using $B_L \approx 35$ nT and estimating the background field value at closest approach as 32.5, the β of the magnetospheric plasma in the vicinity of Callisto is found to be ~ 0.2 .

The closest-approach data set is shown on expanded scale in Fig. 2; several local signatures in the data may be seen. There is a broad region (between 13:25 and 13:40 UT) centred around the closest approach, where the field gradually increased by ~ 7 nT and then returned to the background value. Field rotations lasting typically between three and five minutes occurred between 13:40 and 13:52, when the spacecraft was outbound from Callisto. Because the field enhancement signature was centred around the closest approach to Callisto, we believe that it was caused by Callisto. Fluctuations between 13:40 and 13:52, on the other hand, need not be directly related to the interaction of Callisto with the magnetospheric plasma. These field rotations could have resulted from field-aligned currents which flow in the plasma sheet boundary layer at these distances. Alternatively, some or all of the field rotations may be associated with the plasma generation/loading processes that have enhanced the plasma density in the vicinity of Callisto⁷. We point out that there is no indication of a bow shock in the vicinity of Callisto even though it is possible that the flow is transonic¹².

In order to understand further the Callisto-associated signature, we have plotted in Fig. 3 the field vectors along the trajectory of Galileo. The magnetic perturbation is unlike the signatures observed when Galileo passed by Io^{1,2} and Ganymede³, two other galilean moons of Jupiter recently encountered. The Io encounter and the first encounter with Ganymede also took Galileo through the plasma wakes. At Io, the field rotated little but decreased by 40% in magnitude. The Io signature has been interpreted as resulting from an internal dipolar field^{1,2} or alternatively from Alfvén wing type currents flowing in the vicinity of Galileo's trajectory¹³. During the first close pass of Ganymede, the total field increased by a factor of five near the closest approach and the perturbation vectors were found to point towards a common dipole axis. The Ganymede perturbation has been unambiguously shown to result from an internal dipole source³ which is approximately antiparallel to Ganymede's spin vector and aligned with the averaged background field of Jupiter. The Callisto-associated signature is dissimilar from both Io- and Ganymede-induced signatures. If the perturbation were caused by a dipole moment approximately antiparallel to the averaged background field of Jupiter at Callisto's orbit (and therefore aligned with the spin axis), the perturbation would have been most pronounced in the B_θ (see Fig. 1 legend) component whereas the observed enhancement was primarily in B_r . If the signature were caused by an Io-like interaction between a conducting object and a flowing plasma, a decrease rather than an increase in the total field strength should have occurred in the wake region¹⁴. We can therefore rule out both a simple conducting plasma-satellite interaction, and an internal dipole oriented along the spin axis of Callisto, as the dominant sources of the Callisto-associated signature.

There are, however, other mechanisms that can account for the Callisto-associated signature. For instance, if the Callisto-obstacle is non-conducting and unmagnetized like the Earth's Moon, it would absorb most of the co-rotating magnetospheric plasma impinging upon it. The magnetic field on the other hand would diffuse right through it. The segments of the magnetic flux tubes emerging from Callisto on the wake side would be devoid of incident magnetospheric plasma and the plasma observed⁷ could be cooler plasma

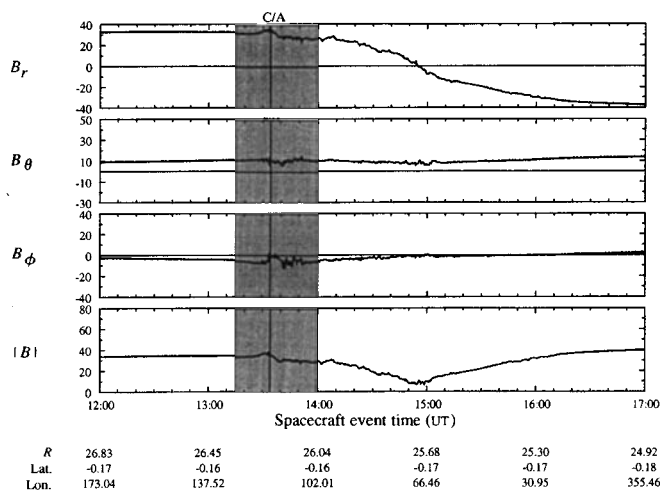


Figure 1 An overview plot of magnetic-field data during the close pass of Callisto on 4 November, 1996. The data are shown in the right-handed System III (epoch 1965) spherical coordinates (r, θ, ϕ)¹⁰. The data in the shaded region were obtained at full resolution ($\Delta t = 0.33$ s) whereas the rest of the data were obtained at a resolution of 24 s. C/A, closest approach.

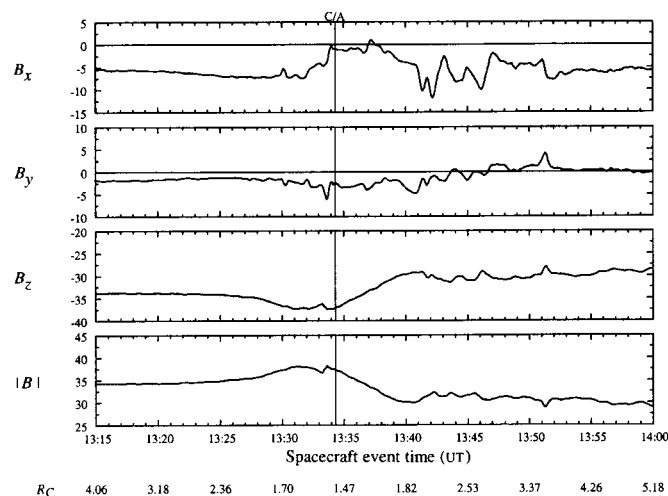


Figure 2 An expanded-scale view of the high-resolution part of the data from Fig. 1 obtained in the close vicinity of Callisto. The data were first averaged by passing them through a 20-s running average window and the resolution was then lowered to 2 s by retaining every ninth point in the time series. The data are in a cartesian coordinate system for which $\hat{x}$ points in the direction of plasma flow, $\hat{y} = \mathbf{B}_0 \times \hat{x} / |\mathbf{B}_0 \times \hat{x}|$ where $\mathbf{B}_0$ is the estimated background field at the closest approach and $\hat{z}$ completes the triad. In this coordinate system, the background magnetic field lies completely in the xz plane.

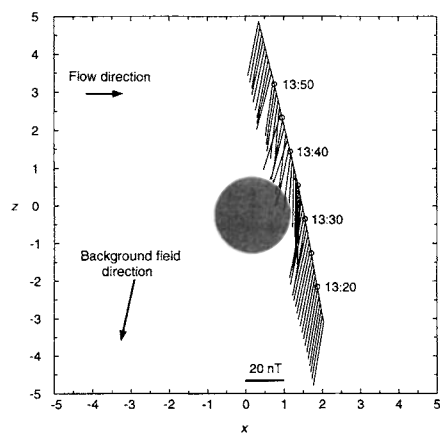


Figure 3 The projection of the trajectory of Galileo during the Callisto fly-by and the magnetic field vectors (time resolution 1 min) in the xz plane. The special coordinate system used is defined in Fig. 2 legend and is different from the one used by Gurnett *et al.* in ref. 7.

from Callisto. The decreased particle pressure would be compensated for by increased magnetic pressure requiring increased field strength in the wake. This is the mechanism invoked to explain the observed diamagnetic cavity in the wake of the Earth's Moon^{15,16} where the enhancement of the field is thought to be produced by currents flowing on the boundaries of the wake region^{16,17}. If Callisto is non-conducting, it would not produce a bow shock even if the plasma flow were super-magnetosonic.

Alternatively, the small enhancement in the field could be attributed to an internal field. If that were the case, the axis of the dipole would have to be tilted by 80° in a direction pointing approximately towards Jupiter, and the surface field at the magnetic equator would be ≤ 14 nT. It seems unlikely that an internal dipole would possess such an orientation. Thus, we believe that the effective dipole perturbation is induced by the interaction of Callisto with the magnetized plasma as described above. In the unlikely event that the signature is that of an internal field, the assumption of no plasma contribution to the wake signal gives an upper bound on the magnetic dipole moment of 2×10^{18} A m² (2×10^{11} T m³). For comparison, the inferred dipole moment for Io^{1,2} is $\leq 4 \times 10^{20}$ A m² and for Ganymede³ is 1.4×10^{20} A m². If a dipole moment is sought which is approximately anti-aligned with the spin axis as has been inferred for Io^{1,2} and Ganymede³, the dipole moment would have to be an order of magnitude smaller. Such a dipole moment would produce a B_θ perturbation < 1 nT near the closest approach and would be barely detectable above the background fluctuations.

Galileo's magnetic field observations therefore show that Callisto is unmagnetized and possibly has a non-conducting exterior. This discovery is plausible in view of the report by Anderson *et al.*¹⁸ that Callisto is probably completely undifferentiated and therefore devoid of a convecting core that could support the generation of an internal magnetic field.

Received 16 December 1996; accepted 26 March 1997.

- Kivelson, M. G. *et al.* A magnetic signature near Io: Initial report from the Galileo magnetometer. *Science* **273**, 337–340 (1996).
- Kivelson, M. G. *et al.* Io's interaction with the plasma torus: Galileo magnetometer report. *Science* **274**, 396–398 (1996).
- Kivelson, M. G. *et al.* Discovery of Ganymede's magnetic field by the Galileo spacecraft. *Nature* **384**, 537–541 (1996).
- Gurnett, D. A., Kurth, W. S., Roux, A., Bolton, S. J. & Kennel, C. F. Evidence for a magnetosphere at Ganymede from plasma-wave observations by the Galileo spacecraft. *Nature* **384**, 535–537 (1996).
- Anderson, J. D., Lau, E. L., Sjogren, W. L., Schubert, G. & Moore, W. B. Gravitational constraints on the internal structure of Ganymede. *Nature* **384**, 541–543 (1996).
- Schubert, G., Zhang, K., Kivelson, M. G. & Anderson, J. D. The magnetic field and internal structure of Ganymede. *Nature* **384**, 544–545 (1996).

- Gurnett, D. A., Kurth, W. S., Roux, A. & Bolton, S. J. Absence of a magnetic-field signature in plasma-wave observations at Callisto. *Nature* **387**, 261–262 (1997).
- Kivelson, M. G., Khurana, K. K., Means, J. D., Russell, C. T. & Snare, R. C. The Galileo magnetic field investigation. *Space Sci. Rev.* **60**, 357–383 (1992).
- Becher, J. W. in *The Physics of the Jovian Magnetosphere* 68–105 (Cambridge Univ. Press, 1983).
- Dessler, A. J. in *The Physics of the Jovian Magnetosphere* 498–504 (Cambridge Univ. Press, 1983).
- Vasyliunas, V. M. in *The Physics of the Jovian Magnetosphere* 395–453 (Cambridge Univ. Press, 1983).
- Barbosa, D. D. On the injection and scattering of protons in Jupiter's magnetosphere. *J. Geophys. Res.* **86**, 8981–8990 (1981).
- Frank, L. A. *et al.* Plasma observations at Io with the Galileo spacecraft. *Science* **274**, 394–395 (1996).
- Neubauer, F. M. Nonlinear standing Alfvén wave current system at Io: Theory. *J. Geophys. Res.* **85**, 1171 (1980).
- Whang, Y. C. Interaction of a magnetized solar wind with the moon. *Phys. Fluids* **11**, 969–975 (1968).
- Colburn, D. S. *et al.* Diamagnetic solar wind cavity discovered behind the moon. *Science* **158**, 1040–1042 (1967).
- Owen, C. J. *et al.* The lunar wake at 6.8 R_L : WIND magnetic field observations. *Geophys. Res. Lett.* **23**, 1263–1266 (1996).
- Anderson, J. D., Lau, E. L., Sjogren, W. L., Schubert, G. & Moore, W. B. Gravitational evidence for an undifferentiated Callisto. *Nature* **387**, 264–266 (1997).

Acknowledgements. We thank S. Joy and A. Frederick for support with programming and plotting, and F. Bagenal for providing information on the background field and plasma conditions in the vicinity of Callisto.

Correspondence should be addressed to K.K.K. (e-mail: kkhurana@igpp.ucla.edu).

Gravitational evidence for an undifferentiated Callisto

J. D. Anderson*, E. L. Lau*, W. L. Sjogren*, G. Schubert† & W. B. Moore†

* Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109-8099, USA

† Department of Earth and Space Sciences, Institute of Geophysics and Planetary Physics, University of California Los Angeles, California 90095-1567, USA

Before the arrival of the Galileo spacecraft at Jupiter, models for the interior structure of the four galilean satellites—Io, Europa, Ganymede and Callisto—ranged from uniform mixtures of rock and ice (that is, undifferentiated objects) or rocky cores surrounded by a mantle of water ice¹. Now it appears that Io has a large metallic core² and that Ganymede is strongly differentiated, most probably into a three-layer structure consisting of a metallic core, a silicate mantle and a deep outer layer of ice³. Direct information on the interior structure of Callisto determined from previous spacecraft fly-bys^{4–6} was essentially limited to an estimate of the mean density being intermediate between pure ice and pure rock. Here we report measurements of Callisto's gravitational field which reveal that, in contrast to Io and Ganymede, this galilean satellite is most probably a homogeneous object consisting of a solar mixture of 40% compressed ice and 60% rock (including iron and iron sulphide). Callisto's undifferentiated state is consistent with the apparent lack of an intrinsic magnetic field⁷, and indicates that the outermost galilean satellite has not experienced a heating phase sufficiently high to separate its rock and metal components from the lighter ices.

Galileo's first encounter with Callisto took place on 4 November 1996, and the gravity signal was detected in the radio Doppler data recorded during the encounter. The data analysis was accomplished by fitting a parametrized orbital model to the radio Doppler data by weighted nonlinear least squares^{8–10}. Callisto's external gravitational field was modelled by the standard spherical harmonic representation of the gravitational potential¹¹. Because we assumed that the orientation of Callisto's principal axes is known, only two non-zero coefficients (J_2 and C_{22}) were included in the model. All other harmonic coefficients were assumed exactly equal to zero. The two included coefficients measure the contributions to the gravitational potential of the spherical harmonics of degree l and order m for $l = 2$, $m = 0$ and $l = 2$, $m = 2$, respectively. In terms of spherical coordinates fixed in the body of Callisto (radius r , latitude ϕ and longitude λ , where r is the radial distance from the centre of mass and longitude is measured from the Callisto–Jupiter line in an

equatorial system defined by Callisto's spin axis), the gravitational potential is

$$V = \frac{GM}{r} \left[1 - \frac{1}{2} J_2 \left(\frac{R}{r} \right)^2 (3 \sin^2 \phi - 1) + 3 C_{22} \left(\frac{R}{r} \right)^2 \cos^2 \phi \cos 2\lambda \right] \quad (1)$$

where G is the gravitational constant and M and R are the mass and radius of Callisto. The closest approach to Callisto occurred at an altitude of 1,136 km, at $\phi = 13.195^\circ$ and $\lambda = 77.917^\circ$ (west longitude).

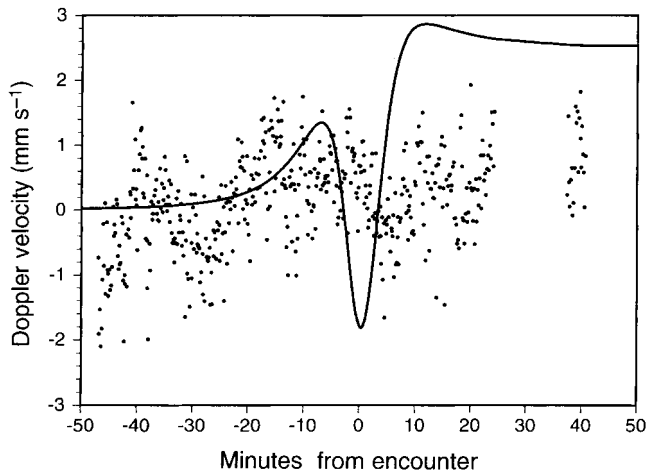


Figure 1 Best fit second-degree gravity signal ($J_2 = 47.7 \times 10^{-6}$ and $C_{22} = 14.3 \times 10^{-6}$) detected in the radio Doppler data from Galileo at Callisto encounter. The Doppler velocity is defined by $c\Delta\nu/\nu$, where $\Delta\nu$ is the Doppler frequency shift referenced to the spacecraft's crystal oscillator, ν is the transmitted frequency, ~ 2.3 GHz, and c is the speed of light. During encounter, the data were recorded by NASA's Deep Space Network tracking facility in Spain. Doppler residuals with respect to the best-fit gravity model are indicated by filled circles. Their standard error is ~ 0.73 mm s $^{-1}$. Outside the displayed encounter interval, the spacecraft transmission was locked to a radio signal from the ground and the standard error is about a factor of four less. Data included in the fit extend from 30 October 1996, 14:52:30 UTC to ~ 7 h and 39 min after closest approach on 4 November 1996, 14:21:00 UTC. The gap in the residuals centred around 30 min from encounter reflects a gap in the raw Doppler data. Because of an unknown bias and drift in the spacecraft's crystal oscillator, the encounter data were assigned a standard error of 1.3 mm s $^{-1}$ (1.8σ) for the error analysis. Any remaining systematic trend in the residuals can be attributed to the propagation of the radio wave through interplanetary plasma.

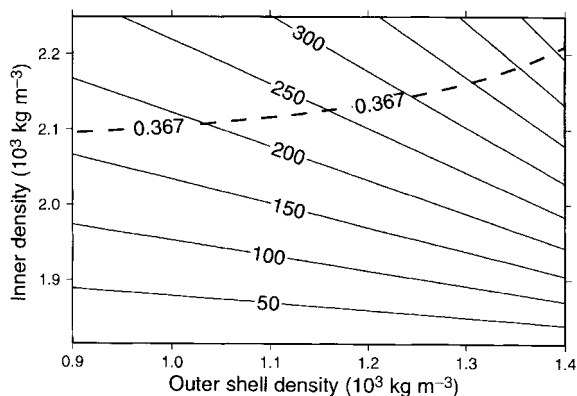


Figure 2 Two-layer model structures of Callisto consistent with its average density and moment of inertia inferred from the observed C_{22} . Contours of ice thickness in kilometres (solid lines) slope downward to the right; the single dashed contour of C/MR^2 , representative of the 1σ lower bound, slopes upward to the right. Values of C/MR^2 larger than 0.4, even though within a standard deviation of the mean, imply a decreasing density with depth and are excluded (see text).

The determination of GM is biased because of a high correlation with Callisto's ephemeris. We are unable to improve on the value ($7,181 \pm 3$ km 3 s $^{-2}$) determined from Pioneer and Voyager encounters⁶. The remaining two gravity coefficients J_2 and C_{22} are not highly correlated with other parameters in the model, but they are highly correlated with each other, so we impose the *a priori* hydrostatic constraint that J_2 is exactly 10/3 of C_{22} . The values that produce the best fit to the encounter Doppler data are $J_2 = (47.7 \pm 11.5) \times 10^{-6}$ and $C_{22} = (14.3 \pm 3.2) \times 10^{-6}$. The Doppler signal generated by these values is shown in Fig. 1. After the fit, the correlation coefficient between J_2 and C_{22} is $\mu = 0.9128$. It may appear from Fig. 1 that the signal-to-noise ratio is not good enough to infer an unambiguous measurement of C_{22} . To the contrary, the data analysis, including an assessment of the covariance matrix from the nonlinear least squares process, leads us to conclude that the errors on the two parameters represent a best estimate of realistic standard error (essentially a 22% accuracy for the second-degree external gravitational field).

Because a homogeneous Callisto is unlikely to have high rigidity, the assumption of hydrostatic equilibrium is plausible. However, unlike our previous results from the Galileo spacecraft's two encounters with Ganymede³, we are not able to place experimental limits on the hydrostatic assumption. In comparison, the value of C_{22} for the Moon (which is clearly non-hydrostatic) is comparable to what would be expected for a highly differentiated Callisto, whereas the Moon's appropriately scaled J_2 is somewhat larger than what we have measured for a homogeneous Callisto¹². If Callisto's measured J_2 and C_{22} were significantly less than what we have found, it could indeed be the result of rigidity lowering the values from their hydrostatic homogeneous values. But we have measured an extreme upper limit to J_2 and C_{22} . Rigidity could hide some degree of differentiation in Callisto if its present distortion is a fossil tidal/rotational bulge from a time in Callisto's history when it was rotating more rapidly. However, it is unlikely that Callisto could ever have rotated fast enough to offset the effect of significant mass concentration toward its centre. It is also difficult to explain how, once Callisto differentiated, its ice mantle or rock core could become rigid enough to preserve the excess fossil bulge of a rapidly spinning Callisto over geological time. For purposes of interior modelling we therefore assume that the measured J_2 and C_{22} reflect a tidal and rotational distortion of Callisto into a fluid equilibrium figure with the present rotation rate of the satellite.

The theory of equilibrium figures for synchronously rotating satellites is well known^{12–15}, and we have applied it previously to Io and Ganymede^{3,3}. By applying the same theory to Callisto, we find that the measured value of C_{22} implies an axial moment of inertia C , normalized to MR^2 , of $C/MR^2 = 0.406$, with 1σ lower bound of 0.367. The axial moment of inertia provides a direct constraint on the internal mass distribution^{16,17}. For a uniform-density body, $C/MR^2 = 0.4$. The smaller the value of C/MR^2 , the more concentrated is the body's mass towards its centre. The determined value of C/MR^2 for Callisto clearly indicates a body of uniform density. Even the 1σ lower bound implies a nearly homogeneous body, given the mean density and radius of Callisto (Fig. 2). An outer ice layer can be at most ~ 300 km thick for $C/MR^2 = 0.367$. But for these models the interior density is between about 1,900 and 2,200 kg m $^{-3}$ (Fig. 2), indicating that the interior would still contain a large fraction of ice, and Callisto's rock/metal and ice would be only partially separated. Apparently, Callisto was never heated enough to substantially separate its ice and rock/metal components to form a liquid metallic core, and generate a magnetic field by dynamo action, as Ganymede has done¹⁸. Callisto is only slightly smaller and less massive than Ganymede, hence accretional and radiogenic energy sources would have heated Callisto only slightly less than Ganymede¹. Ganymede might have received a tidal frictional heat pulse in the past, passing through a resonance with Io and Europa¹⁹, that was large enough to cause a core differentiation and/or initiate core dynamo action. Callisto

would not have experienced such tidal heating and, accordingly, Callisto's failure to differentiate may be attributed to its orbital dynamical isolation as the galilean satellite farthest from Jupiter. □

Received 12 December 1996; accepted 9 April 1997.

1. Schubert, G., Spohn, T. & Reynolds, R. T. in *Satellites* (eds Burns, J. A. & Matthews, M. S.) 629–688 (Univ. Arizona Press, Tucson, 1986).
2. Anderson, J. D., Sjogren, W. L. & Schubert, G. *Science* **272**, 709–712 (1996).
3. Anderson, J. D., Lau, E. L., Sjogren, W. L., Schubert, G. & Moore, W. B. *Nature* **384**, 541–543 (1996).
4. Null, G. W., Anderson, J. D. & Wong, S. K. *Science* **12**, 476–477 (1975).
5. Null, G. W. *Astron. J.* **81**, 1153–1161 (1976).
6. Campbell, J. K. & Synnott, S. P. *Astron. J.* **90**, 364–372 (1985).
7. Khurana, K. K., Kivelson, M. G., Russell, C. T., Walker, R. J. & Southwood, D. J. Absence of an internal magnetic field at Callisto. *Nature* **387**, 262–264 (1997).
8. Moyer, T. D. *Orbit Determination Program* (Tech. Rep. No. TR32–1527, Jet Propulsion Lab., Pasadena, 1971).
9. Anderson, J. D. in *Experimental Gravitation* (ed. Bertotti, B.) 163–199 (Academic, New York, 1974).
10. Tapley, B. D. in *Recent Advances in Dynamical Astronomy* (eds Tapley, B. D. & Szebehely, V.) 396–425 (Reidel, Dordrecht, 1973).
11. Kaula, W. M. *Theory of Satellite Geodesy* (Blaisdell, Waltham, MA, 1966).
12. Hubbard, W. B. & Anderson, J. D. *Icarus* **33**, 336–341 (1978).
13. Kaula, W. M. *An Introduction to Planetary Physics: The Terrestrial Planets* (Wiley, New York, 1968).
14. Dermott, S. F. *Icarus* **37**, 310–321 (1979).
15. Zharkov, V. N., Leontiev, V. V. & Kozenko, A. V. *Icarus* **61**, 92–100 (1985).
16. Jeffreys, H. *The Earth* (Cambridge Univ. Press, 1962).
17. Schubert, G., Limonadi, D., Anderson, J. D., Campbell, J. K. & Giampieri, G. *Icarus* **111**, 433–440 (1994).
18. Schubert, G., Zhang, K., Kivelson, M. G. & Anderson, J. D. *Nature* **384**, 544–545 (1996).
19. Malhotra, R. *Icarus* **94**, 399–412 (1991).

Acknowledgements. This work was sponsored by the Galileo Project and was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA. G.S. and W.B.M. were supported by NASA through the Galileo Project at JPL and the Planetary Geology and Geophysics Program.

Correspondence should be addressed to J.D.A. (e-mail: john.d.anderson@jpl.nasa.gov).

Large low-field magnetoresistance in $\text{La}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$ induced by artificial grain boundaries

N. D. Mathur*, G. Burnell*, S. P. Isaac*, T. J. Jackson*, B.-S. Teo*, J. L. MacManus-Driscoll†, L. F. Cohen‡, J. E. Evetts* & M. G. Blamire*

* Department of Materials Science, University of Cambridge, Cambridge CB2 3QZ, UK

† Department of Materials, Imperial College, London SW7 2BZ, UK

‡ Blackett Laboratory, Imperial College, London SW7 2BZ, UK

A number of different compounds, such as those derived from LaMnO_3 , have recently been shown to exhibit very large changes (up to 10%) in electrical resistance when a magnetic field is applied^{1–4}—a phenomenon known as colossal magnetoresistance (CMR). But magnetic fields of several tesla are typically required to obtain such a large magnetoresistive effect, thus limiting the potential for applications. Nevertheless the complex and intimate link between magnetic structure, crystallographic structure and electrical resistivity in CMR materials, in addition to being of fundamental scientific interest, appears to provide some scope for engineering a more sensitive magnetoresistive response. Here we elucidate the effect of specific structural defects on the CMR behaviour of the compound $\text{La}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$. We have made thin film devices that isolate the contribution of a single grain boundary that was introduced into an epitaxial film of the material by growing it on a bicrystal substrate. These devices display sharp resistance switching in magnetic fields orders of magnitude less than those normally associated with CMR. These results both provide insight into the role of grain boundaries, and demonstrate the potential for developing sub-micrometre magnetic field sensors based on the CMR effect.

In doped perovskite manganites of the form $\text{R}_{1-x}\text{A}_x\text{MnO}_3$, where R and A are rare-earth and alkaline-earth elements respectively, a

marked polarization of charge carriers due to strong intra-atomic Hund interactions on the manganese sites leads to a dependence of the electrical resistivity on the magnetic microstructure^{5–12} as suggested by the ‘double exchange’ model. When the manganese spins are aligned, either ferromagnetically below the Curie temperature T_c , or at higher temperatures in sufficiently high magnetic fields, conduction is metallic and resistivity increases with increasing temperature. Alternatively, in the spin-disordered state, charge conduction is by thermally activated site-to-site hopping and carrier mobility is affected by structural or magnetic distortions (polarons). In this ‘activated transport’ regime the resistivity is high and decreases with increasing temperature. It is thus in the vicinity of T_c that CMR effects are most pronounced, but local disruption of the magnetic order at any temperature has the potential to alter the electrical conduction and induce a magnetoresistive effect. Such a disruption is likely to arise at structural defects. Investigations into the role of defects in CMR materials have hitherto been based on indirect comparisons between single-crystal and polycrystalline samples^{12–14}, or the study of thin resistive barriers in heteroepitaxial structures^{15,16}. A full understanding of the role of defects, and CMR in general, has remained elusive.

To provide insight into the role of grain boundaries and to isolate their contribution to the magnetoresistance, we have studied devices patterned from films grown on bicrystal substrates. Stimulated by the success of investigations into the nature of grain boundaries deliberately introduced into high-temperature superconductors by growth on bicrystal substrates^{17,18}, we have used a SrTiO_3 bicrystal substrate to introduce a single grain boundary into epitaxial thin films of $\text{La}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$ (LCMO). We observe large low-field magnetoresistance across these artificial grain boundaries below T_c .

Epitaxial LCMO (002) films 200 nm thick were grown from a stoichiometric target by pulsed laser deposition (KrF laser, wavelength $\lambda = 248$ nm) on bicrystal substrates consisting of two SrTiO_3 (001) crystals misaligned by 24° at a growth temperature of about 800°C and in an oxygen ambient pressure of 15 Pa. The films were annealed *in situ* at about the same temperature in 500 kPa O_2 for 1 hour. X-ray diffraction measurements confirmed a single (002)

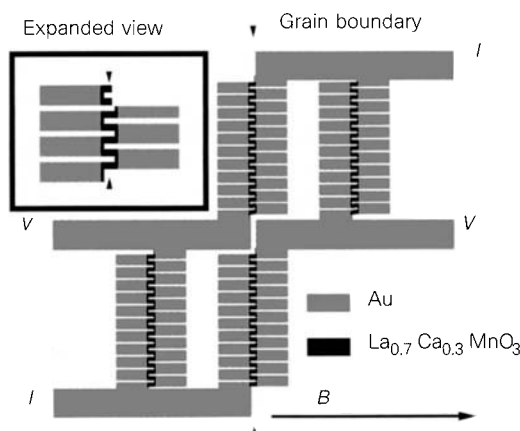


Figure 1 The meander track Wheatstone-bridge geometry used for the grain boundary measurements. Each arm consists of 19 CMR elements, $2\mu\text{m}$ wide (shown black), those on two of the arms crossing the grain boundary (dashed line), the other two sampling control material. The gold (marked in grey) was added to the structure to reduce the resistance of the wiring between the arms and the links between each element. The bridge structure for measurement along the grain boundary was similar except that the meander was replaced by a single straight track in each case.

orientation in each film with a typical rocking-curve width of 0.1° . Texture scans also confirmed an in-plane angular misorientation of 24° between film halves. Films from bicrystal deposition runs showed both a sharp drop in electrical resistivity and the onset of ferromagnetism near 240 K.

To enable us to study the artificial grain boundary directly, we patterned a variety of highly symmetric Wheatstone-bridge structures into the bicrystal films using optical lithography and argon ion milling. The symmetry of the bridge structures ensured that all resistance contributions balanced to zero except those arising from the grain boundary. To lower the resistance of inactive regions and provide contact pads for aluminium wire bonding, 60–70 nm of gold was sputter deposited on each chip through a lift-off mask. We provided bridge structures on each chip in which current was able to flow either across the grain boundary using a meander track (Fig. 1) or along it. The devices were characterized in a liquid nitrogen cryostat in a magnetic field of up to 300 mT by passing a constant current through each bridge and measuring the voltage across the output terminals. This voltage represents a direct measure of the resistance introduced by the grain boundary. As described below, plots of the output voltage were obtained for different values of magnetic field and temperature for two bridge types.

We found that in zero applied field, the resistance of the type of bridge in which current is forced to meander across the grain boundary increased substantially as the temperature was reduced through 240 K (Fig. 2A). This is precisely the opposite of the behaviour that we observed in single-crystal films from the same deposition runs. A large bridge magnetoresistance (27% at 77 K) was observed during magnetic field sweeps within ± 200 mT over a range of temperatures down to 77 K. At low temperatures, a strong low-field hysteresis in bridge resistance was observed (Fig. 3), along with a higher zero field resistance following a zero-field cool (also shown in Fig. 3). At increased temperatures, the magnitude of the peaks in negative magnetoresistance, and the magnitude of the fields at which they occur, both decreased until the peaks disappeared altogether at temperatures around 230 K (Fig. 2B). The peak fields are close to the coercive fields we have measured in epitaxial

samples of the same material. Measurements of an alternative Wheatstone-bridge structure, in which the current direction was along the grain boundary, showed a sharp drop in bridge output near T_c similar to the drop in resistance shown by epitaxial films.

The disappearance of the low-field magnetoresistance above a temperature of about T_c implies a strong link with the ferromagnetic state. The very large magnetoresistive effects which were originally observed in doped perovskites^{1,2,19–21} have only been seen at temperatures near T_c using fields of several tesla and are thus attributed to the quenching of local thermal spin fluctuations. We conclude that the disruption that we have induced in the crystalline order of the material has been sufficient to induce a local spin disorder which is responsible for the observed additional resistance in zero field.

The hysteresis that we have observed (Fig. 3a) indicates that magnetization reversal involves domain wall motion, and is reminiscent of the hysteresis exhibited by metallic multilayers²² which show spin-polarized transport phenomena²³. Given the 24° misalignment of the magnetic 'easy' axes across the grain boundary in each of our bicrystal films, the complete alignment of the magnetic vectors in film halves can only occur when an applied field overcomes any magnetic anisotropies that are present. Resistance changes tend to saturate above 200 mT (Fig. 3a) and identifying this field as a measure of the magnetic anisotropy field suggests an anisotropy constant of order 10^5 J m⁻³. These conclusions, drawn from our study of precisely misaligned epitaxial lattices, are consistent with the work of Hwang *et al.*¹² on polycrystalline samples where the negative magnetoresistance is inferred to arise from spin-polarized transport across magnetic domain boundaries coincident with grain boundaries.

At high temperatures, the hysteresis disappears, and the grain boundary resistance actually rises with increasing field (Fig. 3b). This behaviour arises in the activated transport regime, which dominates above T_c , and would have been masked in previous experiments^{12–16} by the unbalanced contribution of defect-free material. We have thus demonstrated the existence of a positive

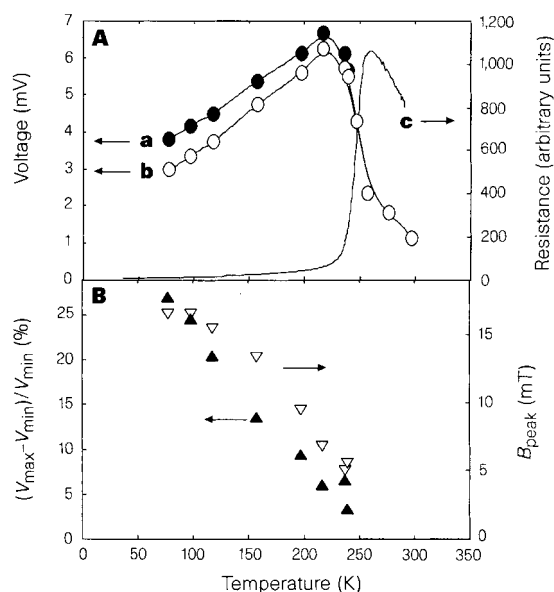


Figure 2 Wheatstone-bridge measurements with $I = 1.5 \mu\text{A}$ with B in-plane and perpendicular to the grain boundary. **A**, Temperature variation of the peak output voltage (trace **a**) and the output voltage at 180 mT (trace **b**). Trace **c** represents the zero field resistance of an unpatterned LCMO film ($I = 0.1$ mA). **B**, Temperature variation of the maximum negative magnetoresistance measured and peak positions which are symmetric about $B = 0$ and occur at $\pm B_{\text{peak}}$. Voltages $V_{\max}$ and $V_{\min}$ are the maximum and minimum output voltages measured.

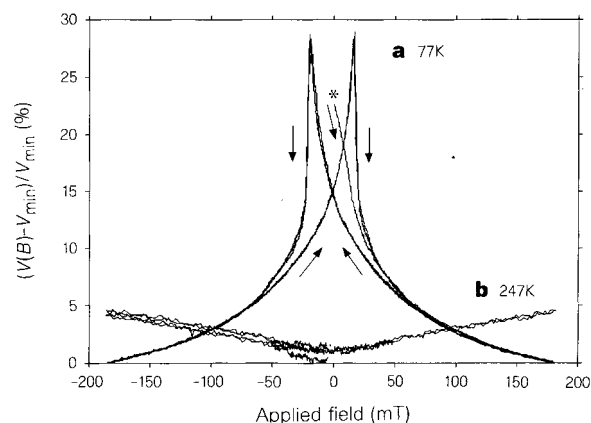


Figure 3 Wheatstone-bridge measurements with $I = 1.5 \mu\text{A}$ with B in-plane and perpendicular to the grain boundary, at **(a)** 77 K and **(b)** 247 K. Voltages $V(B)$ and $V_{\min}$ are the field-dependent output voltage and the minimum output voltage measured. The asterisk indicates data taken following zero-field cooling to 77 K. Data taken following cooling to 77 K in 300 mT are also presented but can barely be distinguished.

magnetoresistive component in CMR materials arising at grain boundaries at high temperatures.

Our highly symmetric Wheatstone-bridge structures afford new insight into the underlying mechanisms that contribute to CMR phenomena. The technique is able to isolate the specific contribution of a single grain boundary and quantify the role of the imposed magnetic substructure. In addition, the discovery of a positive magnetoresistance at higher temperatures underlines the significance of the cross-over between transport regimes near T_c . The dramatic effect induced by the introduction of a single grain boundary demonstrates the extreme sensitivity of the electrical conduction processes in CMR materials to microstructural defects, and the results should stimulate further experimental and theoretical study of these materials.

The potential applications of CMR materials have been limited up to now by the very high magnetic fields required to induce significant changes in resistance. Our findings immediately provide a method by which controllable low-field magnetoresistive devices may be simply fabricated from CMR materials. Moreover, such devices would have the potential to operate beyond the narrow range of operating temperatures of more simply configured CMR material systems and at lower absolute resistivities. The fields required to induce the low-field magnetoresistance in the grain boundary devices are comparable to those required for existing magnetoresistive structures such as spin valves. The use of high- T_c materials (such as $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$) should permit room-temperature operation, and a greater understanding of the basic physics and improvements in materials properties are likely to increase the magnetoresistance achievable. □

Received 4 February; accepted 25 March 1997.

1. Jin, S. *et al.* Thousandfold change in resistivity in magnetoresistive La–Ca–Mn–O films. *Science* **264**, 413–415 (1994).
2. von Helmholt, R., Wecker, J., Holszapfel, B., Schultz, L. & Samwer, K. Giant negative magnetoresistance in perovskite-like $\text{La}_{2/3}\text{Ba}_{1/3}\text{MnO}_3$ ferromagnetic films. *Phys. Rev. Lett.* **71**, 2331–2334 (1993).
3. Subramanian, M. A. *et al.* Colossal magnetoresistance without $\text{Mn}^{4+}/\text{Mn}^{3+}$ double exchange in the stoichiometric pyrochlore $\text{Tl}_2\text{Mn}_2\text{O}_7$. *Science* **273**, 81–84 (1996).
4. Ramirez, A. P., Cava, R. J. & Krajewski, J. Colossal magnetoresistance in Cr-based chalcogenide spinels. *Nature* **386**, 156–159 (1997).
5. Jonker, G. H. & van Santen, J. H. Ferromagnetic compounds of manganese with perovskite structure. *Physica* **16**, 337–349 (1950).
6. Zener, C. Interaction between the d -shells in the transition metals. II. Ferromagnetic compounds of manganese with perovskite structure. *Phys. Rev.* **82**, 403–405 (1951).
7. Anderson, P. W. & Hasegawa, H. Considerations on double exchange. *Phys. Rev.* **100**, 675–681 (1955).
8. de Gennes, P.-G. Effects of double exchange in magnetic crystals. *Phys. Rev. B* **118**, 141–154 (1960).
9. Furukawa, N. Transport properties of the Kondo lattice model in the limit $S = \infty$ and $D = \infty$. *J. Phys. Soc. Jpn* **63**, 3214–3217 (1994).
10. Hwang, H. Y., Cheong, S.-W., Radaelli, P. G., Marezio, M. & Batlogg, B. Lattice effects on the magnetoresistance in doped LaMnO_3 . *Phys. Rev. Lett.* **75**, 914–917 (1995).
11. Hundley, M. F. *et al.* Transport-magnetism correlations in the ferromagnetic oxide $\text{La}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$. *Appl. Phys. Lett.* **67**, 860–862 (1995).
12. Hwang, H. Y., Cheong, S.-W., Ong, N. P. & Batlogg, B. Spin-polarised intergrain tunnelling in $\text{La}_{2/3}\text{Sr}_{1/3}\text{MnO}_3$. *Phys. Rev. Lett.* **77**, 2041–2044 (1996).
13. Gupta, A. *et al.* Grain-boundary effects on the magnetoresistance properties of perovskite manganite films. *Phys. Rev. B* **54**, R15629–R15632 (1996).
14. Snyder, G. J., Hiskes, R., DiCarolis, S., Beasley, M. R. & Geballe, T. H. Intrinsic electrical transport and magnetic properties of $\text{La}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$ and $\text{La}_{0.6}\text{Sr}_{0.4}\text{MnO}_3$ MOCVD thin films and bulk material. *Phys. Rev. B* **53**, 14434–14444 (1996).
15. Lu, Y. *et al.* Large magnetotunnelling effect at low magnetic fields in micrometer-scale epitaxial $\text{La}_{0.6}\text{Sr}_{0.4}\text{MnO}_3$ tunnel junctions. *Phys. Rev. B* **54**, R8357–R8360 (1996).
16. Sun, J. Z. *et al.* Observation of large low-field magnetoresistance in trilayer perpendicular transport devices made using doped manganite perovskites. *Appl. Phys. Lett.* **69**, 3266–3268 (1996).
17. Dimos, D., Chaudhari, P. & Mannhart, J. Superconducting transport properties of grain boundaries in $\text{YBa}_2\text{Cu}_3\text{O}_7$ bicrystals. *Phys. Rev. B* **41**, 4038–4049 (1990).
18. Beck, A. *et al.* Fabrication and superconducting transport properties of bicrystal grain boundary Josephson junctions on different substrates. *IEEE Trans. Appl. Supercond.* **5**, 2192–2195 (1995).
19. McCormack, M. *et al.* Very large magnetoresistance in perovskite-like La–Ca–Mn–O thin films. *Appl. Phys. Lett.* **64**, 3045–3047 (1994).
20. Sun, J. Z., Krusin-Elbaum, L., Parkin, S. S. P. & Xiao, G. Transport and magnetic properties of *in situ* grown thin-film La–Y–Ca–Mn–O. *Appl. Phys. Lett.* **67**, 2726–2728 (1995).
21. Xiong, G. *et al.* Giant magnetoresistance in epitaxial $\text{Nd}_{0.5}\text{Sr}_{0.5}\text{MnO}_3$ thin films. *Appl. Phys. Lett.* **66**, 1427–1429 (1995).
22. Parkin, S. S. P., Bhadra, R. & Roche, K. P. Oscillatory magnetic exchange coupling through thin copper layers. *Phys. Rev. Lett.* **66**, 2152–2155 (1991).
23. Dieny, B. Classical-theory of giant magnetoresistance in spin-valve multilayers—influence of thicknesses, number of periods, bulk and interfacial spin-dependent scattering. *J. Phys. Condens. Matter* **4**, 8009–8020 (1992).

Acknowledgements. We thank Z. H. Barber, W. E. Booij, M. J. Hall, S. Kovtonyuk, D. B. Jardine, D.-J. Kang, E. J. Tarte, Y. S. Soman and R. E. Somekh. This work was supported by the UK EPSRC through the Advanced Magnetism Programme.

Correspondence and requests for materials should be addressed to N.D.M. (e-mail: ndm12@cus.cam.ac.uk).

Synthesis of epothilones A and B in solid and solution phase

K. C. Nicolaou*, N. Winssinger*, J. Pastor*, S. Ninkovic*, F. Sarabia*, Y. He*, D. Vourloumis*, Z. Yang*, T. Li, P. Giannakakou† & E. Hamel‡

* Department of Chemistry and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA, and Department of Chemistry and Biochemistry, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA

† Medicine Branch, DCS, NCI, National Institutes of Health, Bethesda, Maryland 20892, USA

‡ Laboratory of Drug Discovery Research and Development, Developmental Therapeutics Program, DCTDC, NCI, Frederick Cancer Research and Development Center, Frederick, Maryland 21702, USA

Epothilones A and B, two compounds that have been recently isolated¹ from myxobacterium *Sorangium cellulosum* strain 90, have generated intense interest^{2–16} among chemists, biologists and clinicians owing to the structural complexity, unusual mechanism of interaction with microtubules and anticancer potential of these molecules. Like taxol* (refs 17, 18), they exhibit cytotoxicity against tumour cells by inducing microtubule assembly and stabilization^{3,4}, even in taxol-resistant cell lines. Following the structural elucidation of these molecules by X-ray crystallography in 1996¹, several syntheses of epothilones A (refs 12–16) and B (ref. 19) have been reported, indicative of the potential importance of these molecules in the cancer field. Here we report the first solid-phase synthesis of epothilone A, the total synthesis of epothilone B, and the generation of a small epothilone library. The solid-phase synthesis applied here to epothilone A could open up new possibilities in natural-product synthesis and, together with solution-phase synthesis of other epothilones, paves the way for the generation of large combinatorial libraries of these important molecules for biological screening.

The strategy for the solid-phase synthesis of epothilone A (1) was based on the retrosynthetic analysis indicated in Fig. 1 (refs 5, 13). Thus, it was anticipated that the three requisite fragments (5–7), one on a solid support (7), would be coupled together sequentially through an aldol reaction, an esterification reaction, and an olefin metathesis reaction^{20–25}, the latter simultaneously cyclizing and liberating the product from the solid support ($6 + 7 + 5 \rightarrow 4 \rightarrow 3$). A simple desilylation and epoxidation reaction would then complete the total synthesis of epothilone A (1) and analogues thereof ($3 \rightarrow 1$). The outlook for obtaining two products at each of the aldol, metathesis and epoxidation steps was considered advantageous for the purposes of library generation.

Merrifield resin (8, Fig. 2) was converted to phosphonium salt 9 in >90% yield by sequential reaction with: (1) 1,4-butanediol–NaH–*n*-Bu₄NI cat.; (2) Ph₃P–iodine–imidazole; and (3) Ph₃P (for abbreviations see figure legends). Ylide 10²⁶, generated from 9 by the action of NaHMDs in THF:DMSO at 25 °C, reacted with aldehyde 11 (K.C.N. *et al.*, unpublished results) at 0 °C to form olefinic compound 12 in >70% yield. The geometry of the double bond in 12 was tentatively¹⁴ assigned as *Z*, but its geometry was neither rigorously determined nor did it matter for our purposes. Desilylation of 12 with HF–pyridine, followed by Swern oxidation of the resulting primary alcohol furnished aldehyde 7 in high yield (>95%). The aldol condensation of the polymer-bound aldehyde 7 with the dianion derived from keto acid 6 in the presence of ZnCl₂ in tetrahydrofuran (THF) gave a mixture of diastereoisomers

* Bristol-Myers Squibb has registered Taxol as a trademark and wishes the scientific community to use the name paclitaxel.

Figure 1 Retrosynthetic analysis of epothilone A (**1**) by a solid-phase olefin metathesis strategy. TBS, *t*-BuMe₂Si; the shaded circle indicates polystyrene.

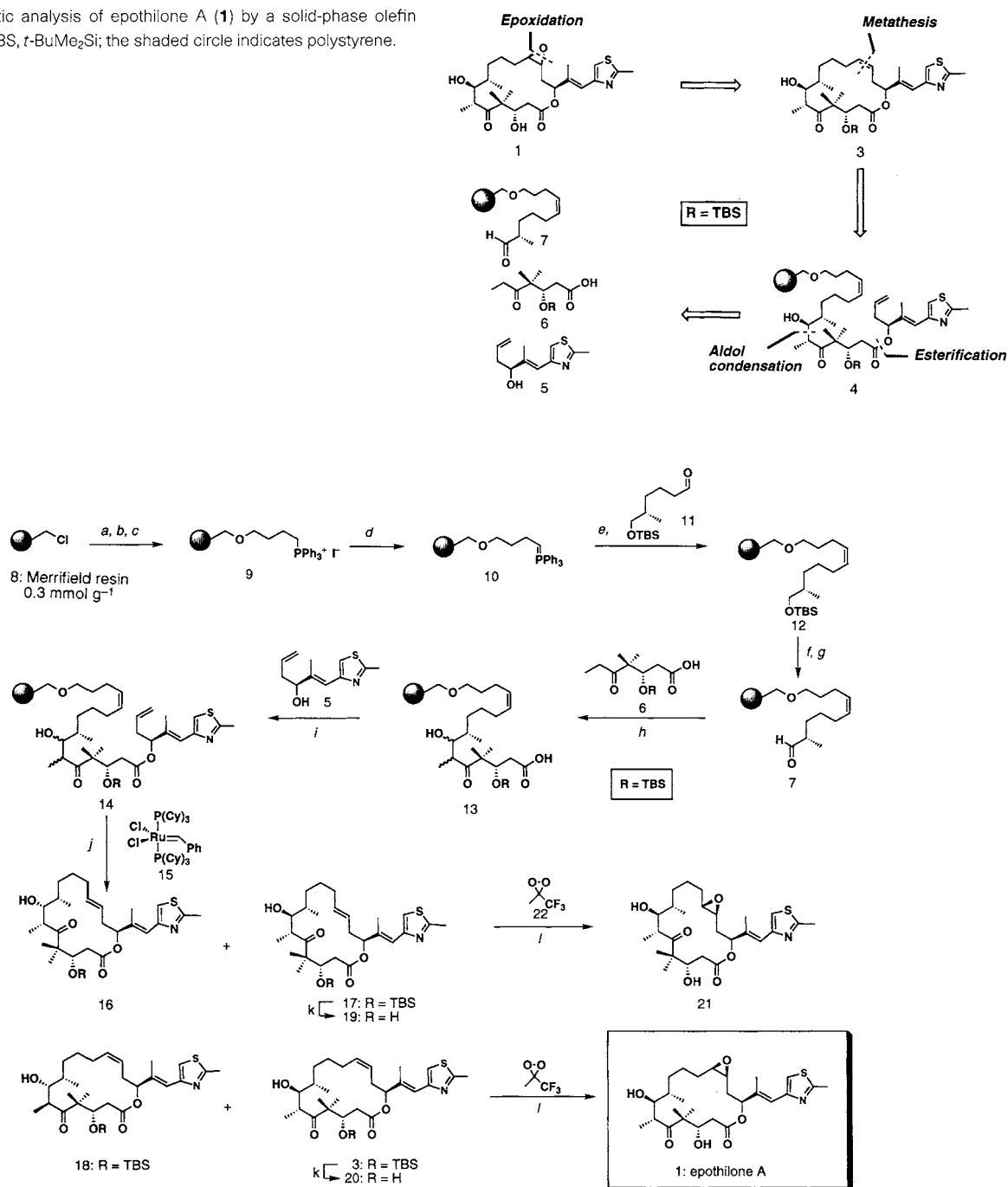


Figure 2 (a) 1,4-butanediol (5.0 equiv.), NaH (5.0 equiv.), *n*-Bu₄Ni (0.1 equiv.), DMF, 25 °C, 12 h; (b) Ph₃P (4.0 equiv.), I₂ (4.0 equiv.), imidazole (4.0 equiv.), CH₂Cl₂, 25 °C, 3 h; (c) Ph₃P (10 equiv.), 90 °C, 12 h (>90% for 3 steps based on mass gain of polymer); (d) NaHMDS (3.0 equiv.), THF : DMSO (1 : 1), 25 °C, 12 h; (e) **11** (2.0 equiv.), THF, 0 °C, 3 h (>70% based on aldehyde recovered from ozonolysis); (f) 10% HF-pyridine in THF, 25 °C, 12 h; (g) (COCl)₂ (4.0 equiv.), DMSO (8.0 equiv.), Et₃N (12.5 equiv.), -78 °C → -25 °C (estimated yield ~95% for 2 step; the reaction was monitored by IR analysis of polymer-bound material and by TLC analysis of the products obtained by ozonolysis); (h) **6** (2.0 equiv.), LDA (2.2 equiv.), THF, -78 °C → -40 °C, 1 h; then add resulting enolate to the resin suspended in a ZnCl₂ (2.0 equiv.) solution in THF, -78 °C → -40 °C, 2 h (~90%; estimated yield, as step g); (i) **5**, (5.0 equiv.), DCC (5.0 equiv.), 4-DMAP (5.0 equiv.), 25 °C, 15 h (80% yield as determined by recovered heterocycle fragments obtained by treatment with NaOMe); (j) **15** (0.75 equiv.), CH₂Cl₂, 25 °C, 48 h (52%); **16** : **17** : **18** : **3** = ~3 : 3 : 1 : 3; (k) 20% TFA in CH₂Cl₂ (v/v), 92% for **19** and 90% for **20**; (l) **22** [methyl(trifluoromethyl)dioxirane, acetonitrile], 0 °C, 2 h (70% for **1**, 45% for **21**; in addition to these products, the corresponding α-epoxides were

also obtained). NaHMDS, sodium bis(trimethylsilyl)amide; DMSO, dimethyl sulphoxide; LDA, lithium diisopropylamide; TBS, *t*-BuMe₂Si; 4-DMAP, 4-dimethylaminopyridine; TFA, trifluoroacetic acid. Selected physical data for compound **20**: ¹H NMR (400 MHz, CDCl₃) δ 6.95 (s, 1 H, ArH), 6.59 (s, 1 H, ArCH = C(CH₃)), 5.44 (ddd, *J* = 10.5, 10.5, 4.5 Hz, 1 H, CH = CHCH₂), 5.36 (ddd, *J* = 10.5, 10.5, 5.0 Hz, 1 H, CH = CHCH₂), 5.28 (d, *J* = 9.4 Hz, 1 H, CO₂CH), 4.23 (d, *J* = 11.1 Hz, 1 H, (CH₃)₂CCl-H(OH)), 3.72 (m, CHOH(CHCH₃)), 3.43–3.37 (m, 1 H, OH), 3.14 (q, *J* = 6.7 Hz, 1 H, CH₃CH(C = O)), 3.05 (bs, 1 H, OH), 2.72–2.63 (m, 1 H), 2.69 (s, 3 H, CH₃Ar), 2.48 (dd, *J* = 14.8, 11.3 Hz, 1 H, CH₂COO), 2.33 (dd, *J* = 14.8, 2.0 Hz, 1 H, CH₂COO), 2.30–2.13 (m, 2 H) 2.07 (s, 3 H, ArCH = CCH₃), 2.07–1.98 (m, 1 H), 1.80–1.60 (m, 2 H), 1.32 (s, 3 H, C(CH₃)₂), 1.36–1.13 (m, 3 H), 1.17 (d, *J* = 6.8 Hz, 3 H, CH₃CH(C = O)), 1.06 (s, 3 H, C(CH₃)₂), 0.99 (d, *J* = 7.0 Hz, 3 H, CH₃CHCH₂); ¹³C NMR (150.9 MHz, CDCl₃) δ 220.6, 170.4, 165.0, 151.9, 138.7, 133.4, 125.0, 119.4, 115.8, 78.4, 74.1, 72.3, 53.3, 41.7, 39.2, 38.5, 32.4, 31.7, 27.6, 27.4, 22.7, 19.0, 18.6, 15.9, 15.5, 13.5; infrared (thin film) ν_{max} 3,453, 2,929, 1,733, 1,686, 1,506, 1,464, 1,250, 978 cm⁻¹; [α]_D²⁵ -80.2 (c 1.36, CHCl₃); HRMS (FAB), calc. for C₂₆H₃₉CsNO₅S (M + Cs⁺) 610.1603, found 610.1580.

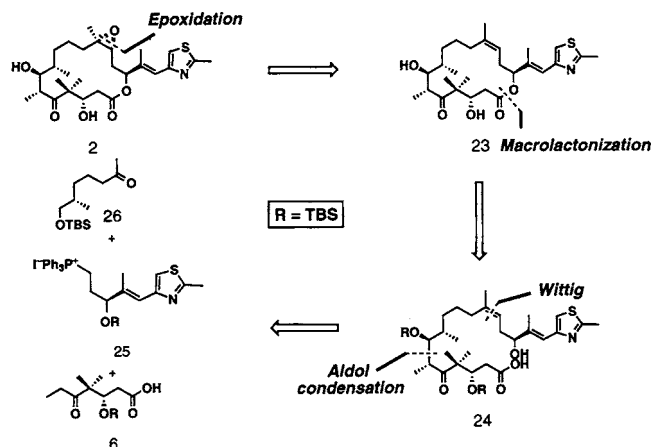


Figure 3 Retrosynthetic analysis of epothilone B (**2**) by a solution-phase strategy. TBS, *t*-BuMe₂Si.

(~90% yield, ~1:1 ratio). Finally, introduction of the heterocyclic segment **5**¹³ onto the growing substrate was achieved by esterification, leading to the required precursor **14** in ~80% yield. Exposure of **14** to RuCl₂(=CHPh)(PCy₃)₂, where Cy is cyclohexyl, catalyst (**15**) in CH₂Cl₂ at 25 °C released from the resin olefinic compounds **16**–**18** and **3** (52% total yield, **16**:**17**:**18**:**3** ≈ 3:3:1:3 as determined by high pressure liquid chromatography (HPLC)). Compounds **16**–**18** and **3** could be separated either by HPLC or by preparative layer silica gel chromatography, and the two with the correct C6–C7 stereochemistry (that is, **17** and **3**) were desilylated by exposure to TFA (see Fig. 2 legend) to afford epothilone precursors **19** (92%) and **20** (90%), respectively. Epoxidation of **19** and **20** with methyl(trifluoromethyl)dioxirane²⁷ then furnished epothilone A (**1**, 70%) and its diastereoisomer **21** (45%), respectively. The α-epoxy isomers of **1** and **21** were also obtained in these epoxidation reactions. Pure synthetic epothilone A (**1**) exhibited identical properties (as determined by thin layer chromatography, [α]_D (optical rotation), ¹H and ¹³C NMR, infrared and HRMS (high resolution mass spectrum)) to those of an authentic sample.

The total synthesis of epothilone B (**2**) followed a strategy derived from the retrosynthetic analysis shown in Fig. 3¹⁴. This strategy called for coupling of intermediates **6**, **25** and **26** via a Wittig olefination, an aldol reaction, and a macrolactonization, followed by epoxidation, and was expected to proceed via intermediates **24** and **23**. This plan was deliberately chosen for its potential to deliver

both diastereoisomers at C6–C7 (aldol reaction) and both geometrical isomers at C12–C13 (Wittig reaction) for molecular diversity and biological screening purposes.

The phosphonium salt **25** (K.C.N. *et al.*, unpublished results) was converted to the corresponding ylide by treatment with NaHMDS which reacted with ketone **26** (K.C.N. *et al.*, unpublished results) to afford a mixture of *Z*- and *E*-olefins **27** in 73% yield and ~1:1 ratio (by ¹H NMR) (Fig. 4). The primary hydroxyl group in **27** was selectively liberated by exposure to CSA (97% yield) and oxidized with SO₃·pyridine–Et₃N–DMSO to afford aldehyde **28** in 95% yield. Treatment of keto acid **6**¹³ with excess LDA in THF, followed by reaction with aldehyde **28**, furnished a mixture of four compounds corresponding to the two geometrical isomers of the C12–C13 double bond and the two diastereomeric isomers at C6–C7 in high yield. This mixture was persilylated by exposure to excess TBSOTf and 2,6-lutidine, and then selectively deprotected at the carboxylic acid site (K₂CO₃–MeOH) to afford chromatographically separable (silica gel) carboxylic acids **29** (31% yield from **28**) and its 6*S*,7*R*-diastereoisomer **29a** (30% yield from **28**).

The tris(silyl ether) **29** was then selectively desilylated at C15 (TBAF, 75% yield) to produce hydroxy acid **24** as a mixture of 12*Z*- and 12*E*-isomers. Macrolactonization¹⁴ of **24** by the Yamaguchi method (2,4,6-trichlorobenzoylchloride, Et₃N, 4-DMAP) resulted in the formation of macrocyclic olefins **30** (40%) and **31** (37%), which were chromatographically separated (silica gel). Exposure of **30** and **31** to TFA led to dihydroxy lactones **32** (89%) and **23** (91%), respectively. Finally, epoxidation of **23** with methyl(trifluoromethyl)dioxirane²⁷ furnished epothilone B (**2**) together with its α-epoxide epimer **35** in 85% yield and ~5:1 ratio in favour of **2**. Pure synthetic epothilone B (**2**) was obtained by preparative layer silica gel chromatography (*R*_f = 0.24, 4% MeOH in CH₂Cl₂) and exhibited identical properties (thin layer chromatography, [α]_D, ¹H and ¹³C NMR, infrared and HRMS) to those of an authentic sample of epothilone B (**2**). Similar treatment of **32** resulted in the formation of epothilones **33** and **34** in 86% yield and ~4:1 ratio. The use of *m*CPBA for these epoxidations gave slightly different results leading to **2** and **35** in 66% total yield and ~5:1 ratio, and **33** and **34** in 73% total yield and ~4:1 ratio.

The synthesized epothilones were tested for their action on tubulin assembly using purified tubulin with an assay²⁸ developed to amplify differences between compounds more active than taxol. As demonstrated in Fig. 5, both epothilone B (**2**) (EC₅₀ = 4.0 ± 1 μM; defined in Fig. 5 legend) and its progenitor **23** (EC₅₀ = 3.3 ± 0.2 μM) were significantly more active than taxol (EC₅₀ = 15.0 ± 2 μM) and epothilone A (**1**) (EC₅₀ =

Table 1 Relative activities of epothilones A (**1**) and B (**2**) as compared with synthetic analogues **23**, **20**, **32**, **34** and taxol

Compound	Induction of tubulin assembly*	Parental	Inhibition of human ovarian carcinoma cell growth†		
			Taxol-resistant		MDR-line
			β-tubulin mutants		
		1A9	1A9PTX10	1A9PTX22	A2780AD
	EC ₅₀ (μM) ± s.d.		IC ₅₀ nM (relative resistance)		
1	14 ± 0.4	2.0	19 (9.5)	4.2 (2.1)	2.4 (1.2)
2	4.0 ± 0.1	0.040	0.035 (0.88)	0.045 (1.1)	0.040 (1.0)
23	3.3 ± 0.2	2.0	33 (17)	3.5 (1.8)	1.5 (0.80)
20	25 ± 1	25	>100 (>4)	75 (3.0)	22 (0.88)
32	39 ± 2	48	>100 (>2)	75 (1.6)	24 (0.50)
34	22 ± 0.9	3.5	30 (8.6)	5.5 (1.6)	3.0 (0.86)
Taxol	15 ± 2	2.0	50 (25)	43 (22)	>100 (>50)

* See Fig. 5.

† The growth of all cell lines was evaluated by quantitation of the protein in microtitre plates⁴. The parental cell line 1A9, a clone of the A2780 cell line, was used to select two taxol resistant sublines (1A9PTX10 and 1A9PTX22)²⁹. These sublines were selected by growth in the presence of taxol and verapamil, a P-glycoprotein modulator. Two distinct point mutations in the β-tubulin isotype M40 gene were identified. In 1A9PTX10 amino acid residue 270 was changed from Phe (TTT) to Val (GTT), and in 1A9PTX22 residue 364 was changed from Ala (GCA) to Thr (ACA). The A2780AD line is a multi-drug resistant (MDR) line expressing high levels of P-glycoprotein³⁰. Relative resistance refers to the ratio of the IC₅₀ value obtained with a resistant cell line to that obtained with the parental cell line.

14.0 $\pm$ 0.4 μ M), whereas compounds **34**, **20** and **32** were less effective than taxol.

Preliminary cytotoxicity experiments with 1A9, 1A9PTX10 (β -tubulin mutant)²⁹, 1A9PTX22 (β -tubulin mutant)²⁹ and A2780AD cell lines revealed a number of interesting results (Table 1). Despite its high potency in the tubulin assembly assay, compound **23** did not display the potent cytotoxicity of **2** against 1A9 cells, being similar to **1** and taxol. These data suggest that whereas the C12-C13 epoxide is

not required for the epothilone-tubulin interaction, it may play an important role in localizing the agent to its target within the cell. Like the naturally occurring epothilones **1** and **2**, analogue **23** showed significant activity against the MDR line A2780AD and the altered β -tubulin-expressing cell lines 1A9PTX10 and 1A9PTX22, suggesting, perhaps, different contact points for the epothilones and taxol with tubulin (that is, stronger binding of epothilones around residue 364 than around 270 relative to taxoids).

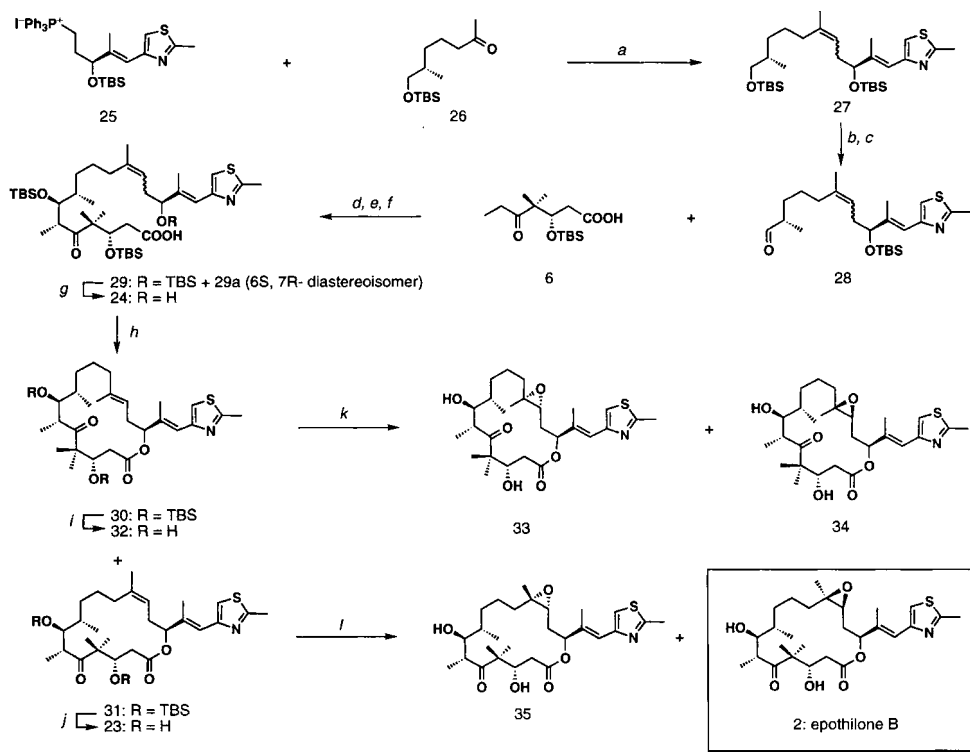


Figure 4 (a) **25** (1.5 equiv.), NaHMDS (1.5 equiv.), THF, 0°C, 15 min; then add **26** (1.0 equiv.), -20°C, 12 h, 73% (*Z*:*E* = 1:1); (b) CSA (1.0 equiv.), CH₂Cl₂:MeOH (1:1), 0°C, 1 h; then 25°C, 0.5 h, 97%; (c) SO₃-pyr. (2.0 equiv.), DMSO (10 equiv.), Et₃N (5 equiv.), CH₂Cl₂, 25°C, 0.5 h, 95%; (d) LDA (3.0 equiv.), THF, 0°C, 15 min; then **6** (1.2 equiv. in THF), -78 $\rightarrow$ -40°C, 0.5 h; then **28** (1.0 equiv. in THF), -78°C; (e) TBSOTf (3.0 equiv.), 2,6-lutidine (5.0 equiv.), CH₂Cl₂, 0°C, 2 h; (f) K₂CO₃ (2.0 equiv.), MeOH, 25°C, 15 min, 31% of **29** from **28** and 30% of **29a** from **28**; (g) TBAF (6.0 equiv.), THF, 25°C, 8 h, 75%; (h) 2,4,6-trichlorobenzoylchloride (2.0 equiv.), Et₃N (2.0 equiv.), THF, 0°C, 1 h; then add to a solution of 4-DMAP (10.0 equiv. in toluene, 0.002 M), 25°C, 12 h, 40% of **30** and 37% of **31**; (i) 20% TFA (by volume) in CH₂Cl₂, -10 $\rightarrow$ -0°C, 1 h, 89%; (j) same as i, 91%; (k) methyl(trifluoromethyl) dioxirane, acetonitrile, 0°C, 86% (**33**:**34** ~1:1 diastereoisomers) or *m*CPBA (1.5 equiv.), benzene, 3°C, 2 h, 73% (**33**:**34** ~4:1 ratio of stereoisomers); (l) methyl(trifluoromethyl)dioxirane, acetonitrile, 0°C, 85% (**2**:**35** ~5:1 ratio of diastereoisomers) or *m*CPBA (1.5 equiv.), benzene, 3°C, 2 h, 66% (**2**:**35** ~5:1 ratio of diastereoisomers); NaHMDS, sodium bis(trimethylsilyl)amide; CSA, camphorsulphonic acid; DMSO, dimethyl sulphoxide; LDA, lithium diisopropylamide; TBS, *t*-

BuMe₂Si; TBSOTf, *t*-BuMe₂SiOSO₂CF₃; TBAF, tetra-*n*-butylammonium fluoride; 4-DMAP, 4-dimethylaminopyridine; *m*CPBA, 3-chloroperoxybenzoic acid; TFA, trifluoroacetic acid. Selected physical data for compound **23**: ¹H NMR (600 MHz, CDCl₃) δ 6.94 (s, 1H, SCH = C), 6.57 (s, 1H, CH = CCH₃), 5.20 (d, *J* = 9.7 Hz, 1H, CH₂COOCH), 5.13 (dd, *J* = 9.6, 4.6 Hz, 1H, CH₂C = CHCH₂), 4.28 (d, *J* = 9.7 Hz, 1H, (CH₂)₂CCHOH), 3.71 (s, 1H, CHOH), 3.47 (bs, 1H, OH), 3.15 (q, *J* = 6.8 Hz, 1H, C(O)CHCH₃), 3.04 (bs, 1H, OH), 2.68 (s, 3H, N = C(CH₃)S), 2.62 (ddd, *J* = 15.0, 10.2, 10.1 Hz, 1H, CH₂CH = CCH₃), 2.45 (dd, *J* = 14.7, 11.1 Hz, 1H, CH₂COOCH), 2.38-2.24 (m, 1H), 2.28 (dd, *J* = 14.8, 2.2 Hz, CH₂COOCH), 2.22 (d, *J* = 14.9 Hz, 1H, CH₂C(CH₃) = CHCH₂), 2.06 (s, 3H, CH = CCH₃), 1.90-1.84 (m, 1H), 1.76-1.69 (m, 1H), 1.65 (s, 3H, CH₂C(CH₃) = CH), 1.33 (s, 3H, C(CH₃)₂), 1.32-1.22 (m, 4H), 1.19 (d, *J* = 6.8 Hz, 3H, CH(CH₃)), 1.06 (s, 3H, C(CH₃)₂), 1.00 (d, *J* = 7.0 Hz, 3H, CH(CH₃)); ¹³C NMR (150.9 MHz, CDCl₃) δ 220.4, 170.2, 164.9, 151.8, 139.1, 138.3, 120.8, 119.1, 115.5, 78.9, 74.1, 72.3, 53.6, 41.7, 39.7, 32.6, 31.8, 31.7, 25.4, 23.0, 19.1, 18.1, 16.0, 15.8, 13.5; infrared (thin film) ν_{max} 3,460, 2,954, 2,919, 1,725, 1,684, 1,455, 1,379, 1,290, 1,249, 1,184, 1,143, 1,043, 1,008, 973, 750 cm⁻¹; [α]_D²⁵ -91.5 s. (c 0.3, CHCl₃); HRMS (FAB) *m/e* 492.2795, (*M* + H⁺) calc. for C₂₇H₄₁NO₅S 492.2784.

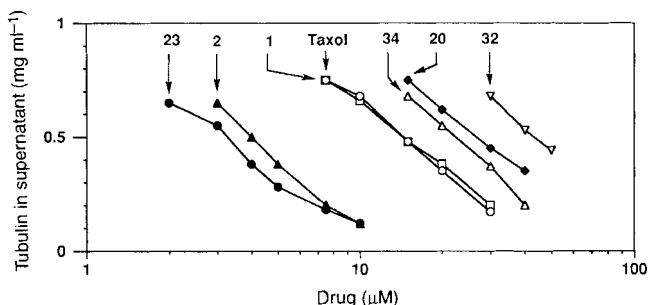


Figure 5 The tubulin assembly assay was performed essentially as described previously²⁸. Reaction mixtures contained purified tubulin at 1.0 mg ml⁻¹, 0.4 M monosodium glutamate, 5% dimethyl sulphoxide, and varying drug concentrations. Each compound was evaluated in three different experiments and average values are shown. The EC₅₀ is defined as the drug concentration that causes 50% of the tubulin to assemble into polymer. In the absence of drug, <5% of the tubulin was removed by centrifugation, while with high concentrations of the most active drugs, >95% of the protein formed polymer. This suggests that at least 90% of the tubulin had the potential to interact with epothilones and taxoids. Although the EC₅₀ value obtained for taxol was higher than that obtained in an alternate assay³, the agent's role in these experiments was only as a control. The numbers on the curves correspond to compound numbers in the text.

The solid-phase synthesis of epothilone A (**1**) described here represents a new concept for the total synthesis of natural products, traces a highly efficient pathway to the naturally occurring epothilones, and opens the way for the generation of large combinatorial epothilone libraries. The biological results demonstrate that more potent microtubule binding analogues than the parent epothilones can be obtained (for example, compound **23**) by chemical synthesis. Furthermore, our findings point to lipophilic substituents rather than the epoxide moiety as important elements for binding activity. The role of the epoxide in the cytotoxicity of epothilones, however, still remains to be elucidated. □

Received 3 April; accepted 17 April 1997.

- Höfle, G. *et al.* Epothilone A and B—novel 16-membered macrolides with cytotoxic activity: isolation, crystal structure, and conformation in solution. *Angew. Chem. Int. Edn. Engl.* **35**, 1567–1569 (1996).
- Grever, M. R., Schepartz, S. A. & Chabner, B. A. The national cancer institute: cancer drug discovery and development program. *Semin. Oncol.* **19**, 622–638 (1992).
- Bollag, D. M. *et al.* Epothilones, a new class of microtubule-stabilizing agents with a Taxol-like mechanism of action. *Cancer Res.* **55**, 2325–2333 (1995).
- Kowalski, R. J., Giannakakou, P. & Hamel, E. Activities of the microtubule-stabilizing agents epothilones A and B with purified tubulin and in cells resistant to paclitaxel (Taxol). *J. Biol. Chem.* **272**, 2534–2541 (1997).
- Nicolaou, K. C., He, Y., Vourloumis, D., Vallberg, H. & Yang, Z. An approach to epothilones based on olefin metathesis. *Angew. Chem. Int. Edn. Engl.* **35**, 2399–2401 (1996).
- Meng, D., Sorensen, E. J., Bertinato, P. & Danishefsky, S. J. Studies toward a synthesis of epothilone A: use of hydropyran templates for the management of acyclic stereochemical relationships. *J. Org. Chem.* **61**, 7998–7999 (1996).
- Bertinato, P., Sorensen, E. J., Meng, D. & Danishefsky, S. J. Studies toward a synthesis of epothilone A: stereocontrolled assembly of the acyl region and models for macrocyclization. *J. Org. Chem.* **61**, 8000–8001 (1996).
- Schinz, D., Limberg, A. & Böhm, O. M. Studies toward the total synthesis of epothilones: asymmetric synthesis of the key fragments. *Chem. Eur. J.* **2**, 1477–1482 (1996).
- Mulzer, J. & Mantoulidis, A. Synthesis of the C(1)–C(9) segment of the cytotoxic macrolides epothilone A and B. *Tetrahedr. Lett.* **37**, 9179–9182 (1996).
- Claus, E., Pahl, A., Jones, P. G., Meyer, H. M. & Kaleske, M. Synthesis of the C1–C9 segment of epothilones. *Tetrahedr. Lett.* **38**, 1359–1362 (1997).
- Gabriel, T. & Wessjohann, L. The Chromium–Reformatsky reaction: asymmetric synthesis of the aldol fragment of the cytotoxic epothilones from 3-(2-bromoacetyl)-2-oxazolidinones. *Tetrahedr. Lett.* **38**, 1363–1366 (1997).
- Balog, A. *et al.* Total synthesis of (–)-epothilone A. *Angew. Chem. Int. Edn. Engl.* **35**, 2801–2803 (1996).
- Yang, Z., He, Y., Vourloumis, D., Vallberg, H. & Nicolaou, K. C. Total synthesis of epothilone A: the olefin metathesis approach. *Angew. Chem. Int. Edn. Engl.* **36**, 166–168 (1997).
- Nicolaou, K. C., Sarabia, F., Ninkovic, S. & Yang, Z. Total synthesis of epothilone A: the macrocyclization approach. *Angew. Chem. Int. Edn. Engl.* **36**, 525–527 (1997).
- Schinz, D., Limberg, A., Bauer, A., Böhm, O. M. & Cordes, M. Total synthesis of (–)-epothilone A. *Angew. Chem. Int. Edn. Engl.* **36**, 523–524 (1997).
- Meng, D. *et al.* Remote effects in macrolide formation through ring-forming olefin metathesis: an application to the synthesis of fully active epothilone congeners. *J. Am. Chem. Soc.* **119**, 2733–2734 (1997).
- Horwitz, S. B., Fant, J. & Schiff, P. B. Promotion of microtubule assembly *in vitro* by taxol. *Nature* **277**, 665–667 (1979).
- Nicolaou, K. C., Dai, W.-M. & Guy, R. K. The chemistry and biology of Taxol. *Angew. Chem. Int. Edn. Engl.* **33**, 15–44 (1994).
- Su, D.-S. *et al.* Total synthesis of (–)-epothilone B: an extension of the Suzuki coupling method and insights into structure–activity relationships of the epothilones. *Angew. Chem. Int. Edn. Engl.* **36**, 757–759 (1997).
- Grubbs, R. H., Miller, S. J. & Fu, G. C. Ring-closing metathesis and related processes in organic synthesis. *Acc. Chem. Res.* **28**, 446–452 (1995).
- Miller, S. J., Blackwell, H. E. & Grubbs, R. H. Application of ring-closing metathesis to the synthesis of rigidified amino acids and peptides. *J. Am. Chem. Soc.* **118**, 9606–9614 (1996).
- Xu, Z., Johannes, C. W., Salzman, S. S. & Hoveyda, A. H. Enantioselective total synthesis of antifungal agent Sch 38516. *J. Am. Chem. Soc.* **118**, 10926–10927 (1996).
- van Maarseveen, J. H. *et al.* Solid phase ring-closing metathesis: cyclization/cleavage approach towards a seven membered cycloolefin. *Tetrahedr. Lett.* **37**, 8249–8252 (1996).
- Miller, J. F., Termin, A. & Piscopio, A. D. Ester enolate Claisen-ring closing metathesis route to functionalized carbocycles and heterocycles. Presented at 213th Am. Chem. Soc. Natl Meeting, San Francisco. 13–17 April (1997).
- Schuster, M., Pernerstorfer, J. & Blechert, S. Ruthenium-catalyzed metathesis of polymer-bound olefins. *Angew. Chem. Int. Edn. Engl.* **35**, 1979–1980 (1996).
- Bernard, M. & Ford, W. T. Wittig reagents bound to cross-linked polystyrenes. *J. Org. Chem.* **48**, 326–332 (1983).
- Yang, D., Wong, M.-K. & Yip, Y.-C. Epoxidation of olefins using methyl(trifluoromethyl)dioxirane generated *in situ*. *J. Org. Chem.* **60**, 3887–3889 (1995).
- Lin, C. M. *et al.* A convenient tubulin-based quantitative assay for paclitaxel (Taxol) derivatives more effective in inducing assembly than the parent compound. *Cancer Chemother. Pharmacol.* **38**, 136–140 (1996).
- Giannakakou, P. *et al.* Paclitaxel resistant human ovarian cancer cells have mutant β -tubulins that exhibit impaired paclitaxel driven polymerization. *J. Biol. Chem.* (in the press).
- Rogan, A. M. *et al.* Reversal of adriamycin resistance by verapamil in human ovarian cancer. *Science* **244**, 994–996 (1984).

Acknowledgements We thank G. Höfle and Merck Research Laboratories for gifts of epothilones A and B; and D. H. Huang and G. Siuzdak for NMR and mass spectroscopic assistance, respectively. This work was supported by the NIH, The Skaggs Institute for Chemical Biology and the CaP CURE Foundation, and fellowships from the Ministerio de Educación y Ciencia (Spain) (J.P.), Fundación Ramón Areces (Spain) (F.S.) and Novartis (D.V.), and grants from Merck, DuPont-Merck, Schering Plough, Hoffmann La Roche and Amgen.

Correspondence and requests for materials should be addressed to K.C.N. at Scripps Research Institute.

Evolution of the nitrogen cycle and its influence on the biological sequestration of CO₂ in the ocean

Paul G. Falkowski

Oceanographic and Atmospheric Sciences Division, Brookhaven National Laboratory, Upton, New York 11973, USA

Over geological time, photosynthetic carbon fixation in the oceans has exceeded respiratory oxidation of organic carbon. The imbalance between the two processes has resulted in the simultaneous accumulation of oxygen in, and drawdown of carbon dioxide from, the Earth's atmosphere, and the burial of organic carbon in marine sediments^{1–3}. It is generally assumed that these processes are limited by the availability of phosphorus^{4,5}, which is supplied by continental weathering and fluvial discharge^{5–7}. Over the past two million years, decreases in atmospheric carbon dioxide concentrations during glacial periods correlate with increases in the export of organic carbon from surface waters to the marine sediments^{8–11}, but variations in phosphorus fluxes appear to have been too small to account for these changes^{12,13}. Consequently, it has been assumed that total oceanic primary productivity remained relatively constant during glacial-to-interglacial transitions, although the fraction of this productivity exported to the sediments somehow increased during glacial periods^{12,14}. Here I present an analysis of the evolution of biogeochemical cycles which suggests that fixed nitrogen, not phosphorus, limits primary productivity on geological timescales. Small variations in the ratio of nitrogen fixation to denitrification can significantly change atmospheric carbon dioxide concentrations on glacial-to-interglacial timescales. The ratio of these two processes appears to be determined by the oxidation state of the ocean and the supply of trace elements, especially iron.

Globally, nitrogen and phosphorus are the two elements that potentially limit the biologically mediated carbon assimilation in the oceans by photoautotrophs. It is frequently argued that, as N₂ is abundant in both the ocean and atmosphere, and, in principle, can be biologically reduced to the equivalent of NH₃ by N₂-fixing cyanobacteria (that is, diazotrophs), nitrogen cannot be limiting on geological timescales^{4,15,16}. It then follows that phosphorus, which has no significant atmospheric source, must ultimately limit biological productivity. The underlying assumptions of these tenets should, however, be considered within the context of the evolution of biogeochemical cycles and the manifestations of those cycles in the contemporary ocean.

Virtually all fixed inorganic nitrogen in the contemporary ocean is oxidized to nitrate. Where did the nitrate come from? Although, in the Archaean atmosphere, electrical discharge or bolide impacts might have promoted NO formation from reaction between N₂ and CO₂, the yield for the reaction is low¹⁷. NH₃ in the Archaean atmosphere would have photodissociated, driven by ultraviolet radiation¹⁷; however, N₂ would have been stable and abundant^{17,18}. N₂ can be biologically reduced to NH₃ via the enzyme nitrogenase. Biological N₂ fixation is a strictly anaerobic process¹⁹, and the sequence of the genes encoding the catalytic subunits for nitrogenase is highly conserved in cyanobacteria and other eubacteria, strongly suggesting an ancient, common ancestral origin²⁰. The antiquity and homology of nitrogen fixation capacity also implies that fixed inorganic nitrogen in the Archaean and early Proterozoic oceans was scarce before the evolution of diazotrophic organisms; that is, there was strong evolutionary selection for N₂ fixation.

In contrast to fixed inorganic nitrogen, soluble phosphorus (largely as inorganic phosphate), in the Archaean oceans was probably relatively abundant in the mildly reducing environment of the time^{2,21}. Whereas apatite and other calcium-based and substituted solid phases of phosphate minerals precipitated in the primary formation of crustal sediments, secondary reactions of phosphate with Al and transition metals such as Fe are mediated at either low salinity or in sulphate-depleted waters (for example forming vivianite²²) low pH (for example forming strengite), or high oxidation states of the cations²³. Although these reactions would decrease the overall soluble phosphate concentration, the initial condition of the Archaean ocean probably had a very low reactive N:P ratio in the dissolved inorganic phase. As N₂ fixation proceeded, that ratio would have increased with a build-up of ammonium in the ocean interior. The accumulation of fixed nitrogen in the oceans would continue until the N:P ratio of the inorganic elements reached equilibrium with the N:P ratio of the sedimenting particulate organic matter (POM). Presumably, the latter ratio would approximate that of extant, nitrogen-fixing marine cyanobacteria, which is 16:1 by atoms^{4,24} or greater²⁵. In the absence of any substantial loss processes, the accumulation of the fixed inorganic nitrogen would ultimately have been constrained by the availability of phosphate.

The formation of nitrate from ammonium is sequentially catalysed through a nitrite intermediate by two groups of aerobic bacteria: one group oxidizes ammonium to nitrite, the second oxidizes nitrite to nitrate. Both processes require molecular oxygen; hence, nitrification must have evolved after the formation of free O₂ in the oceans by oxygenic photoautotrophs. Nitrification also provides reductant for the chemoautotrophic reduction of inorganic carbon, but is thermodynamically much less efficient than photosynthetic carbon fixation²⁶. Therefore, from a geological perspective, the conversion of ammonium to nitrate probably proceeded rapidly and provided a substrate (namely, NO₃⁻), that eventually could serve both as a source of nitrogen for photoautotrophs and as an electron acceptor for a diverse group of heterotrophic, anaerobic bacteria, the denitrifiers.

In the sequence of the three major biological processes that comprise the nitrogen cycle, denitrification must have been the last to emerge. This process, which permits the reduction of NO₃⁻ to (ultimately) N₂, occurs in the modern ocean in three major regions, namely, continental margin sediments, areas of restricted circulation such as fjords, and oxygen minima zones of perennially stratified seas²⁷⁻³⁰. Denitrification seems to have evolved independently several times; the organisms and enzymes responsible for the pathway are highly diverse, and the pathway is found in both Archaeobacteria and Eubacteria³¹.

With the emergence of denitrification, a circuit for nitrogen was established between the atmosphere and ocean, where N₂ fixation provides a source of nitrogen for photoautotrophic production, but simultaneously, fixed nitrogen is removed from the oceans as N₂. The equilibrium between these two processes is bounded; the ratio of fixed inorganic nitrogen to dissolved inorganic phosphate in the ocean interior can only be less than, or at most equal to, that of the sinking flux of the two elements in POM.

There are three major inferences that can be drawn from the foregoing discussion. First, because the ratio of the sinking flux of particulate organic N and particulate P exceeds the N:P ratio of the dissolved pool of inorganic nutrients in the ocean interior, on average the upward flux of inorganic nutrients must be slightly enriched in P relative to N in relation to the elemental requirements of the photoautotrophs⁴. Hence, although there are some exceptions^{32,33}, at the present time, dissolved, inorganic fixed nitrogen generally limits primary production throughout most of the world oceans^{16,34-36}. Second, the N:P ratio of the dissolved pool of inorganic nutrients in the ocean interior was established by biological processes, not vice versa⁴. The average Redfield N:P ratio of

16:1 for particulate organic matter^{4,24} is an upper bound for the two elements in the dissolved inorganic phase in the ocean interior. In the contemporary ocean, the average N:P ratio of the dissolved inorganic nutrients in the ocean interior is ~14.7 by atoms³⁷, or less³⁸. The deficit in dissolved inorganic fixed nitrogen relative to soluble phosphate in the contemporary ocean implies a slight imbalance between nitrogen fixation and denitrification on time scales of ~10³ to 10⁴ years (ref. 39). Finally, if the net biologically mediated exchange of CO₂ between the atmosphere and ocean requires a change in either the concentration of limiting nutrients or the efficiency of their utilization⁵, and if dissolved inorganic nitrogen rather than phosphate limits productivity in the oceans, then it follows that the ratio of nitrogen fixation to denitrification has a critical role in determining the net biologically mediated exchange of CO₂ between atmosphere and ocean³⁹.

In the oligotrophic open ocean, the major nitrogen-fixing organisms are non-heterocystic cyanobacteria in the genus *Trichodesmium*⁴⁰. That there are no heterocystic marine cyanobacteria and few other planktonic, free-living marine diazotrophs suggests that some factor(s) have limited the abundance and speciation of these organisms in the oceans.

Trichodesmium spp. are obligate diazotrophs⁴¹, hence their distribution depends on the ability to synthesize nitrogenase. Nitrogenase contains two catalytic subunits, and requires Fe and Mo (or, in some organisms, V) to facilitate electron transfer reactions⁴². The catalytic activity per mole of Fe (that is, the rate of enzymatic turnover per mole of metal) is the lowest of any iron-containing enzyme involved in nitrogen metabolism⁴², and the iron requirement for diazotrophic growth is ~100-fold higher than that with fixed nitrogen⁴². The required transition metals, especially Fe, were relatively abundant under the mildly reducing conditions of the Archaean and early Proterozoic oceans^{2,43}, but the photosynthetic production of oxygen by cyanobacteria resulted in the precipitation of Fe(III) complexes, leaving much of the upper ocean with relatively low inventories of bioavailable iron². The loss in bioavailable iron seems to coincide with an evolutionary divergence of cyanobacteria, with the radiation of new marine species that were incapable of nitrogen fixation⁴⁴.

In the modern ocean, aeolian transport of continentally derived minerals is a significant, if not the most important, source of iron to the central ocean basins⁴⁵. In regions far removed from such fluxes, such as the South Pacific, the abundance of *Trichodesmium* seems to be extremely low, whereas in areas with higher fluxes, such as the North Atlantic, Indian Ocean, and the relatively iron-rich regions of the North Pacific, *Trichodesmium* blooms are much more frequent and extensive^{25,46}. The patterns of spatial and temporal distribution of *Trichodesmium* blooms have led to the suggestion that nitrogen fixation in the world oceans is limited by trace metals, especially iron^{25,47,48,68}. Whereas the role of iron in limiting photosynthesis in high-nitrate, low-chlorophyll regions of the world ocean has been established by direct experimental manipulation of iron concentrations⁴⁹⁻⁵¹, the larger role of iron (or other trace elements) in limiting nitrogen fixation in low-nutrient, low-chlorophyll regions of the subtropical ocean gyres might have a far greater impact on the biologically mediated net flux of atmospheric CO₂ into the oceans.

During recent glacial periods, the depression in sea level and corresponding reduction in continental shelf area, combined with a decrease in the intensity of water-column stratification in tropical regions⁵², seems to have resulted in decreased rates of denitrification relative to nitrogen fixation^{53,54}. Consequently, the ratio of dissolved inorganic nitrogen to phosphate could have 'caught up' with that of the sinking flux, effectively increasing the oceanic inventory of fixed inorganic nitrogen and simultaneously enhancing the effect of the biological CO₂ pump. During glacial periods, the net drawdown of atmospheric CO₂ would have been further accelerated by aeolian transport of minerals to the central ocean basins. The higher flux of iron to the oceans would not only have further stimulated the

biological pump by facilitating the biological utilization of pre-formed nutrients in the high-nitrate, low-chlorophyll regions⁵⁵, but, more importantly, would have stimulated N₂ fixation in the low-nutrient, low-chlorophyll regions. Given a C:N ratio of ~6.6 by atoms for the synthesis of new organic matter in the euphotic zone, and an average net deficit of dissolved inorganic nitrogen over phosphate in the modern ocean of 2.7 $\mu\text{mol kg}^{-1}$ (ref. 38), restoration of the N:P ratio to 16:1 in the ocean interior implies that during glacial periods the biological pump could have sequestered an additional 600 Pg C. This amount of carbon is approximately equal to the total global anthropogenic emissions since the beginning of the Industrial Revolution. Direct fertilization of the high-nitrate low-chlorophyll (HNLC) regions of the world oceans with iron would sequester about 140 Pg C (ref. 56). These two fluxes are additive.

Ice core records suggest that atmospheric CO₂ declined from ~290 to 190 $\mu\text{mol mol}^{-1}$ over a period of ~10,000 years during the last interglacial-glacial maximum (~20–100 kyr BP)⁵⁷. An equilibrium, three-box model calculation suggests that 800 Pg C would have to have been fixed by marine photoautotrophs if the ocean's biological pump were to account for the change in atmospheric CO₂. This model accounts for repartitioning of CO₂ between the atmosphere and the upper ocean, including internal adjustments in the equilibrium distributions of the major inorganic carbon species⁵⁸. The calculated change in atmospheric CO₂ would have required an addition of ~3 Tg fixed N per annum in addition to the utilization of nutrients in the HNLC regions. The net flux of fixed N requires only a small change in the global ratio of N₂ fixation to denitrification. Biological N₂ fixation in the contemporary ocean is poorly quantified; estimates range from ~20 Tg per annum for the entire ocean⁵⁹, to >50 Tg for the North Atlantic alone^{60,61}. The simultaneous decrease in denitrification during glacial periods would have further contributed to the net influx of fixed nitrogen in the oceans^{53,54,61}, requiring even less N₂ fixation to achieve the same result.

The enhancement of the effect of the biological pump in sequestering atmospheric CO₂ in the ocean interior by increasing the availability of fixed inorganic nitrogen would have exerted a positive climatic feedback. The initial (for example Milankovich⁶²) forcing would have led to increased cooling if atmospheric CO₂ were removed by a strengthening of the biological CO₂ pump³⁶. This process is self-limiting; when the N:P ratio of the dissolved inorganic nutrients caught up to that of the sinking flux of particulates, phosphorus would become limiting and the biological CO₂ pump would then approach a new steady state. During interglacial periods, however, a reduction in aeolian fluxes of trace elements could lead to a slight decrease in nitrogen fixation relative to denitrification³⁹ and to a slow readjustment towards higher atmospheric CO₂ levels. A nonlinear climate model, incorporating the effect of changes in the oceanic inventory of fixed nitrogen, has been developed that accounts for the termination of glacial cycles via the outgassing of the greenhouse gas, N₂O, as a consequence of increased denitrification in the oceans⁶³.

The concept that nitrogen limits the net effect of the biological CO₂ pump on geological timescales has been proposed previously^{28,36,63}. It has generally been assumed, however, that the increased flux of fixed nitrogen to the oceans during glacial periods was a consequence of continental weathering rather than an enhancement of nitrogen fixation by increased aeolian fluxes of trace elements. The tendency to overlook nitrogen fixation as a cause of variability in the biological pump apparently results from an historical misunderstanding of the nitrogen cycle in the ocean. Redfield⁴ was aware that the N:P ratio of dissolved inorganic nutrients in the ocean interior was lower than that required for photoautotrophic production; however, he concluded that nitrogen could not be limiting because biological N₂ fixation "is so active that there is no difficulty in assuming that it might serve in adjusting the

phosphorus:nitrogen ratio in the sea". That viewpoint was based on an understanding of nutrient cycles in lacustrine environments⁶⁴, where N₂ fixation is generally stimulated when fixed inorganic nitrogen becomes limiting. Redfield did not anticipate the possibility that, in the oceans, but not lakes, N₂ fixation is often limited by trace metals^{47,48}. Ironically, the concept of phosphorus limitation, which is generally embraced by geochemists (see, for example, ref. 65), cannot be supported by the bulk distribution of either dissolved inorganic N or P relative to the sinking flux of the particulate organic forms of these elements in the contemporary ocean, nor is it consistent with an analysis of the evolution of the nitrogen cycle in the ocean or with the sedimentary records that suggest enhanced biological organic fluxes and apparently lower rates of denitrification in low-latitude regions of the oceans during glacial maxima^{53,54,61,66,67}. Although phosphorus concentrations set an upper bound to carbon fixation, that bound is neither geochemically nor ecologically relevant as long as the sinking flux of N:P in organic matter exceeds that of the upwelling flux of inorganic nutrients, regardless of the relative turnover of the two elements. There is no geological evidence that phosphorus has ever actually limited primary production in the world's oceans.

The present, non-steady-state flux of CO₂ in the atmosphere will ultimately lead to a significant redistribution of carbon in the major reservoirs. The feedbacks induced by this potential forcing remain obscure⁶⁸. The processes described that affect the nitrogen cycle, and its potential influence on the biological pump, are complex. They involve the hydrological cycle as it affects the availability of continental sources of trace elements through desertification, thermal contrasts between the ocean and land as they affect aeolian transport vectors of iron or other trace metals, ocean circulation as it affects stratification and fluxes of phosphate and oxygen, as well as biological processes as they affect the transformation and fluxes of the key elements. At present, biogeochemical models do not adequately represent these interconnected processes, and consequently it is difficult to predict even the sign, let alone the magnitude, of the change in the ratio of nitrogen fixation/denitrification in the coming century. None the less, that this ratio has a key role in establishing atmospheric CO₂ concentrations is becoming increasingly clear^{36,39,53,54}, and cannot be ignored in ocean biogeochemical models^{5,12}. □

Received 15 July 1996; accepted 20 March 1997.

1. Berner, R. A. A model for atmospheric CO₂ over the Phanerozoic. *Am. J. Sci.* **291**, 339–376 (1991).
2. Holland, H. D. *The Chemical Evolution of the Atmosphere and Oceans* (Princeton Univ. Press, Princeton, 1984).
3. Walker, J. C. G. Was the Archaean Biosphere upside down? *Nature* **329**, 710–712 (1987).
4. Redfield, A. C. The biological control of chemical factors in the environment. *Am. Sci.* **46**, 205–221 (1958).
5. Broecker, W. S. Glacial to interglacial changes in ocean chemistry. *Prog. Oceanogr.* **11**, 151–197 (1982).
6. Meybeck, M. Carbon, nitrogen, and phosphorus transport by world rivers. *Am. J. Sci.* **282**, 401–450 (1982).
7. Froelich, P. N. Kinetic control of dissolved phosphate in natural rivers and estuaries: a primer on the phosphate buffer mechanism. *Limnol. Oceanogr.* **33**, 649–668 (1988).
8. Mix, A. C. Influence of productivity variations on long-term atmospheric CO₂. *Nature* **337**, 541–544 (1989).
9. Sarinthein, M., Winn, K., Duplessy, J.-C. & Fontugne, M. R. Global variations of surface ocean primary productivity in low and mid latitudes: influence on CO₂ reservoirs of the deep ocean and atmosphere during the last 21,000 years. *Paleoceanography* **3**, 361–399 (1988).
10. Lyle, M. W., Prahl, F. G. & Sparrow, M. A. Upwelling and productivity changes inferred from a temperature record in the central equatorial Pacific. *Nature* **355**, 812–815 (1992).
11. Bender, M. & Sowers, T. The Dole effect and its variations during the last 130,000 years as measured in the Vostok ice core. *Global Biogeochem. Cycles* **8**, 363–376 (1994).
12. Boyle, E. A. The role of vertical chemical fractionation in controlling late quaternary atmospheric carbon dioxide. *J. Geochem. Res.* **93**, 15701–15714 (1988).
13. Imbrie, J. et al. On the structure and origin of major glaciation cycles. I. Linear responses to Milankovitch forcing. *Paleoceanography* **7**, 701–738 (1992).
14. Archer, D. & Maier-Reimer, E. Effect of deep-sea sedimentary calcite preservation on atmospheric CO₂ concentration. *Nature* **367**, 260–263 (1994).
15. Broecker, W. S., Peng, T.-H. & Engh, R. Modeling the carbon system. *Radiocarbon* **22**, 565–598 (1980).
16. Barber, R. T. in *Primary Productivity and Biogeochemical Cycles in the Sea* (eds Falkowski, P. G. & Woodhead, A.) 89–106 (Plenum, New York, 1992).
17. Kasting, J. F. Bolidic impacts and the oxidation state of carbon in the Earth's early atmosphere. *Origins Life Evol. Biosphere* **20**, 199–231 (1990).
18. Warneck, P. *Chemistry of the Natural Atmosphere* (Academic, New York, 1988).
19. Postgate, J. R. *Nitrogen Fixation* (Edward Arnold, London, 1987).
20. Zehr, J. P. et al. Diversity of heterotrophic nitrogen-fixation genes in a marine cyanobacterial mat. *Appl. Environ. Microbiol.* **61**, 2527–2532 (1995).
21. Van Cappellen, P. & Ingall, E. D. Redox stabilization of the atmosphere and oceans by phosphorus-limited marine productivity. *Science* **271**, 493–496 (1996).

Scour in large braided rivers and the recognition of sequence stratigraphic boundaries

James L. Best* & Philip J. Ashworth†

* Department of Earth Sciences, University of Leeds, Leeds LS2 9JT, UK

† School of Geography, University of Leeds, Leeds LS2 9JT, UK

Alluvial scour into shallow marine sediments may be caused by the incision of a river adjusting to a new base level^{1–4} following a fall in sea level. The identification of such erosion surfaces^{1–3} has therefore been pivotal in the reconstruction of past sea-level changes from ancient sedimentary sequences^{1–14}. Here we report data from a study of the Jamuna river, Bangladesh, one of the world's largest modern braided rivers¹⁵, which illustrate that bed scour associated with channel confluences and bends alone can be substantial—as much as five times greater than the mean channel depth. Indeed, the basal erosion surfaces produced by such deep scours have characteristics similar to those of boundaries in some ancient sedimentary sequences that have been assumed to result from sea-level fall^{1–14}, potentially leading to radically different interpretations of past variation in base level and climate. We suggest that, to discount unambiguously the influence of fluvial scour in ancient sediments, the erosive boundary should be greater than five times the mean channel depth and extend for distances greater than the floodplain width. Ideally, it should be traceable between different basins.

Within ancient sedimentary successions, the interpretation of base-level fall, erosive 'sequence' boundaries and the accumulation of sediments in incised valley fills is commonly assessed using a number of diagnostic criteria^{3,5}. These are diachronous juxtaposition across the erosive sequence boundary of sediments deposited in different environments (such as fluvial sediments lying stratigraphically above shallow marine or shoreface deposits)^{1–3,5,7}; erosional relief on the sequence boundary that is significantly greater than the mean fluvial channel depth^{3–9}; a regionally extensive basal erosional surface^{3,7,9}; the presence of 'interfluvial' sediments, such as palaeosols, that are laterally correlative with the sequence boundary and characterize the valley margins^{3,6,7,10–12}; and an erosional boundary that can be correlated between basins^{1–4}, a criterion which is rarely applied^{8–10}.

Interpretation of some sequence boundaries, however, can be difficult where it is unclear what role is played by channel avulsion (the sudden switching of a river's course) and the prevailing alluvial sediment flux¹³, or by contemporary coastal erosion¹⁴. Additionally, a key factor in recognition of sequence boundaries is the depth of 'autocyclic' scour within alluvial channels which owes its origin to intrinsic channel flow and sediment transport processes¹⁶. Although many studies of ancient sediments have inferred sea-level fall when the depth of incision along the sequence boundary is larger than the estimated mean channel depth^{5,6,8}, such guidelines have not been tested critically with scour and channel change data from large¹⁵ modern rivers. Yet these data are essential both to verify the criteria used in the recognition of sequence boundaries and to provide margins of error for the interpretation of sea-level fall. Reliable scour data from braided rivers, such as those detailed here, may also have especial importance for the recognition of sequence boundaries, as the higher gradients of the alluvial plain generated following sea-level fall can favour formation of multichannel rivers^{5,9}.

One of the world's largest river channel confluences is between the sand-bedded Jamuna and Ganges rivers in Bangladesh. These rivers, which have braidplains up to 15 km wide, have a combined

22. Berner, R. A. *Early Diagenesis, a Theoretical Approach* (Princeton Univ. Press, Princeton, 1980).
23. Stumm, W. & Morgan, J. J. *Aquatic Chemistry* (Wiley, New York, 1981).
24. Copin-Montegut, C. & Copin-Montegut, G. Stoichiometry of carbon, nitrogen and phosphorus in marine particulate matter. *Deep-Sea Res.* **30**, 31–46 (1983).
25. Letelier, R. M. & Karl, D. M. The role of *Trichodesmium* spp. in the productivity of the subtropical North Pacific Ocean. *Mar. Ecol. Prog. Ser.* **133**, 263–273 (1996).
26. Kaplan, W. A. in *Nitrogen in the Marine Environment* (eds Carpenter, E. J. & Capone, D. G.) 139–190 (Academic, New York, 1983).
27. Christensen, J. P., Murray, J. W., Devol, A. H. & Codispoti, L. A. Denitrification in continental shelf sediments has major impact on the oceanic nitrogen budget. *Global Biogeochem. Cycles* **1**, 97–116 (1987).
28. Codispoti, L. A. & Christensen, J. P. Nitrification, denitrification and nitrous oxide cycling in the eastern tropical south Pacific Ocean. *Mar. Chem.* **16**, 277–300 (1985).
29. Devol, A. H. Direct measurement of nitrogen gas fluxes from continental shelf sediments. *Nature* **349**, 319–321 (1991).
30. Seitzinger, S. Denitrification in freshwater and coastal marine ecosystems: ecological and geochemical importance. *Limnol. Oceanogr.* **33**, 702–724 (1988).
31. Zumpft, W. G. in *The Prokaryotes: A Handbook on the Biology of Bacteria: Ecophysiology, Isolation, Applications* (eds Ballow, A., Truper, H. G., Dworkin, M., Harder, W. & Schleifer, K.-H.) 554–582 (Springer, New York, 1992).
32. Krom, M. D., Kress, N., Brenner, S. & Gordon, L. I. Phosphorus limitation of primary productivity in the eastern Mediterranean Sea. *Limnol. Oceanogr.* **36**, 424–432 (1991).
33. Martin, J. H. in *Primary Productivity and Biogeochemical Cycles in the Sea* (eds Falkowski, P. & Woodhead, A.) 123–137 (Plenum, New York, 1992).
34. Rhyther, J. H. & Dunstan, W. M. Nitrogen, phosphorus, and eutrophication in the coastal marine environment. *Science* **171**, 1008–10013 (1971).
35. Dugdale, R. C. Nutrient limitation in the sea: dynamics, identification, and significance. *Limnol. Oceanogr.* **12**, 685–695 (1967).
36. McElroy, M. B. Marine biological controls on atmospheric CO₂ and climate. *Nature* **302**, 328–329 (1983).
37. Fanning, K. A. Nutrient provinces in the sea: concentration ratios, reaction rate ratios, and ideal covariation. *J. Geophys. Res.* **97C**, 5693–5712 (1992).
38. Anderson, L. A. & Sarmiento, J. L. Redfield ratios of remineralization determined by nutrient data analysis. *Global Biogeochem. Cycles* **8**, 65–80 (1994).
39. Codispoti, L. A. Is the ocean losing nitrate? *Nature* **376**, 724 (1995).
40. Carpenter, E. J. & Romans, K. Major role of the cyanobacterium *Trichodesmium* in nutrient cycling in the North Atlantic Ocean. *Science* **254**, 1356–1358 (1991).
41. Carpenter, E. J. & McCarthy, J. J. Nitrogen fixation and uptake of combined nitrogenous nutrients by *Oscillatoria* (*Trichodesmium*) *thiebautii* in the western Sargasso Sea. *Limnol. Oceanogr.* **20**, 389–401 (1975).
42. Raven, J. A. The iron and molybdenum use efficiencies of plant growth with different energy, carbon and nitrogen sources. *New Phytol.* **109**, 279–287 (1988).
43. Williams, R. J. P. Natural selection of the elements. *Proc. R. Soc. Lond.* **B213**, 361–397 (1981).
44. Falkowski, P. G. & Raven, J. A. *Aquatic Photosynthesis* (Blackwell, Oxford, 1997).
45. Duce, R. A. & Tindale, N. W. Atmospheric transport of iron and its deposition in the ocean. *Limnol. Oceanogr.* **36**, 1715–1726 (1991).
46. Carpenter, E. J. in *Nitrogen in the Marine Environment* (eds Carpenter, E. J. & Capone, D. G.) (Academic, New York, 1983).
47. Reuter, J. G. J. Theoretical Fe limitations of microbial N₂ fixation in the oceans. *Estuaries* **63**, 445 (1982).
48. Howarth, R. W., Marino, R. & Cole, J. J. Nitrogen fixation in freshwater, estuarine, and marine ecosystems. 2. Biogeochemical controls. *Limnol. Oceanogr.* **33**, 688–701 (1988).
49. Martin, J. H. et al. Testing the iron hypothesis in the Equatorial Pacific. *Nature* **371**, 123–129 (1994).
50. Kolber, Z. S. et al. Iron limitation of phytoplankton photosynthesis in the Equatorial Pacific Ocean. *Nature* **371**, 145–149 (1994).
51. Behrenfeld, M., Bale, A., Kolber, Z., Aiken, J. & Falkowski, P. G. Confirmation of iron limitation of phytoplankton photosynthesis in the equatorial Pacific Ocean. *Nature* **383**, 508–511 (1996).
52. Herguera, J. C. & Berger, W. H. Glacial to postglacial drop in productivity in the western equatorial Pacific: mixing rate vs. nutrient concentration. *Geology* **22**, 629–632 (1994).
53. Ganeshram, R. S., Pedersen, T. F., Calvert, S. E. & Murray, J. W. Large changes in oceanic nutrient inventories from glacial to interglacial periods. *Nature* **376**, 755–758 (1995).
54. Altabet, M. A., Francois, R., Murray, D. W. & Prell, W. L. Climate-related variations in denitrification in the Arabian Sea from sediment ¹⁵N/¹⁴N ratios. *Nature* **373**, 506–509 (1995).
55. Sarmiento, J. & Orr, J. Three-dimensional simulations of the impact of Southern Ocean nutrient depletion on atmospheric CO₂ and ocean chemistry. *Limnol. Oceanogr.* **36**, 1928–1950 (1991).
56. Peng, T.-H. & Broecker, W. S. Factors limiting the reduction of atmospheric CO₂ by iron fertilization. *Limnol. Oceanogr.* **36**, 1919–1927 (1991).
57. Raynaud, D. et al. The ice record of greenhouse gases. *Science* **259**, 926–934 (1993).
58. Dickson, A. G. & Millero, F. W. A comparison of the equilibrium constant for the dissociation of carbonic acid in seawater media. *Deep-Sea Res.* **34**, 1733–1743 (1987).
59. Carpenter, E. J. & Capone, D. G. in *Marine Pelagic Cyanobacteria: Trichodesmium and Other Diazotrophs* (eds Carpenter, E. J., Capone, D. G. & Rueter, J. G.) 211–217 (Academic, Dordrecht, 1992).
60. Gruber, N. & Sarmiento, S. L. Global patterns of marine fixation and denitrification revealed by the conservative tracer N^{*}. *Global Biogeochem. Cycles* (in the press).
61. Farrell, J. W., Pedersen, T. F., Calvert, S. E. & Nielsen, B. Glacial-interglacial changes in nutrient utilization in the equatorial Pacific Ocean. *Nature* **377**, 514–517 (1995).
62. Berger, A. Milankovitch theory and climate. *Rev. Geophys.* **26**, 624–657 (1988).
63. Shaffer, G. A non-linear climate oscillator controlled by biogeochemical cycling in the ocean: an alternative model of Quaternary ice ages cycles. *Clim. Dyn.* **4**, 127–143 (1990).
64. Hutchinson, G. E. *A Treatise on Limnology* (Wiley, New York, 1957).
65. Broecker, W. S. & Peng, T.-H. *Tracers in the Sea* (Eldigio Press, Lamont-Doherty Geological Observatory, New York, 1982).
66. Codispoti, L. A. in *Productivity of the Ocean: Present and Past* (eds Berger, W. H., Smetacek, V. S. & Wefer, G.) 377–394 (Wiley, New York, 1989).
67. Paytan, A., Kastner, M. & Chavez, F. P. Glacial to interglacial fluctuations in productivity in the equatorial Pacific as indicated by marine barite. *Science* **274**, 1355–1357 (1996).
68. Sarmiento, J. L. & Le Quére, C. Oceanic carbon dioxide uptake in a model of century-scale global warming. *Science* **274**, 1346–1350 (1996).
69. Michaels, A. F. et al. Inputs, losses and transformations of nitrogen and phosphorus in the pelagic North Atlantic Ocean. *Biogeochemistry* **35**, 181–226 (1996).

Acknowledgements. I thank the US Dept of Energy and NASA for support, Richard Barber, Philip Boyd, Jürgen Holfort, Andrew Knoll, Tom Pedersen, Jorge Sarmiento, Sybil Seitzinger and Doug Wallace for discussions, and D. Canfield and T. Tyrrell for comments.

Correspondence should be addressed to the author (e-mail: falkowsk@sun2.bnl.gov).

annual mean discharge of $40,000 \text{ m}^3 \text{ s}^{-1}$ and a sediment transport rate of 1.2×10^3 million tons per year (ref. 17), one of the highest rates of any modern river. The basis for our analysis is five bathymetric surveys that were made within the framework of the River Survey Project¹⁷, one of the projects conducted under the Bangladesh Flood Action Plan. The five bathymetric surveys, each covering an area of $10 \times 13 \text{ km}$, were made at the confluence of the Jamuna and Ganges rivers during a 28-month period from October 1993 to January 1996. Survey transects were spaced 200 m apart, generating between 23,000 and 53,000 data points for each bathymetric map. All data were reduced to Standard Low Water level (SLW), an inclined plane that is derived from long-term river-stage records. SLW at Aricha is 2.6 m above present sea level. The discharge of both rivers is dominated by the monsoon hydrograph with non-coincident flood peaks in the Jamuna and Ganges usually occurring in July and late August respectively.

Bathymetric plots from each survey (Fig. 1a–e) show two very deep scours, locally up to 30 m below SLW and five times the mean upstream channel depth (6 m), located at the Jamuna–Ganges confluence. Comparison of these surveys illustrates the remarkable mobility of these scours, which migrate up to 1.8 km during the April–October monsoon flood period (compare Fig. 1a, b with Fig. 1c; Fig. 1d with Fig. 1e). A 15-m-deep scour hole is also present on the outer bend of the western Jamuna anabranch upstream of the junction. Confluence, bend and protrusion scours up to 44 m deep have also been documented in other regions of the Bengali main rivers¹⁸. Maximum scour depths may also be greater during the flood hydrograph as some infilling of the major scours occurs during low flow (compare Fig. 1a with Fig. 1b and Fig. 1c with Fig. 1d). The area of the main confluence scour greater than 20 m below SLW extends for 2 km in length and 0.4 km in width. Changes in bed height between the first and last surveys (Fig. 1f) show

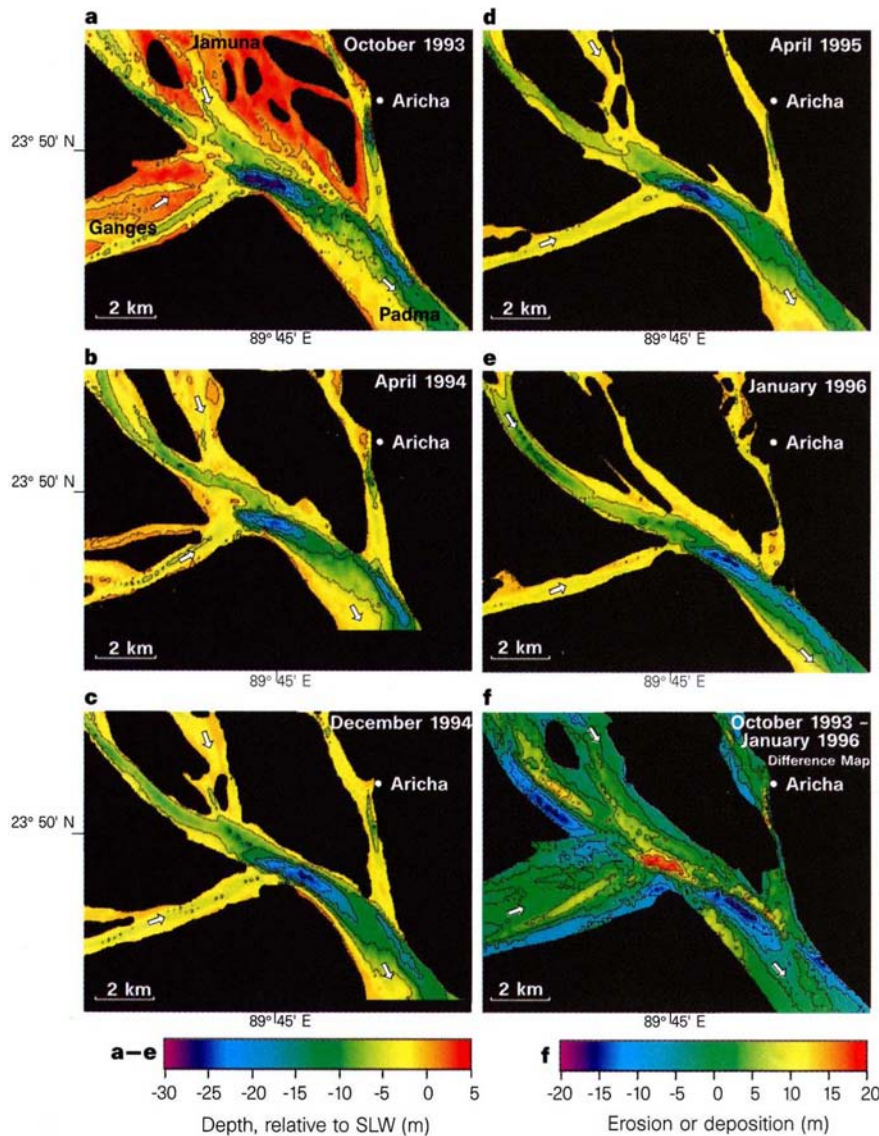


Figure 1 Bathymetry of the Jamuna–Ganges confluence region, Bangladesh, over a 28-month period, for five surveys between October 1993 and January 1996. Bed heights are expressed relative to a Standard Low Water datum (SLW) at Aricha, obtained from long-term hydrological records, and the contour interval is 5 m. Surveys were made from boats using a differential Global Positioning system to fix location to within $\pm 5 \text{ m}$ and echo-sounders with a resolution of $\pm 0.05 \text{ m}$ to obtain bed height. Bar margins and bank edges were defined from the lateral extent of the shallow-draft boat surveys. Interpolation between survey lines is

achieved by kriging (where interpolation uses a variogram model which minimizes the variance of any estimation errors) onto a 25 m grid using a spherical distribution. Parts **b** and **c** have a smaller survey coverage in the Padma River. Parts **a–e** show bed morphology for each survey period with superimposed position of the bar and bank margins. Part **f** shows the change in bed morphology between the October 1993 and January 1996 surveys, illustrating net erosion ($<0 \text{ m}$) and net deposition ($>0 \text{ m}$) between survey periods (again, contour interval is 5 m). Data provided by Delft Hydraulics/Danish Hydraulics Institute¹⁷.

erosion and deposition up to 14 and 17 m respectively. During this time, the Jamuna–Ganges confluence scour shifted down channel by about 3.5 km, whilst the outer bend scour in the western anabranch of the Jamuna migrated laterally by 1 km (Fig. 1f). The deepest regions of these two sites show migration rates of 1.8 and 0.6 km yr⁻¹ for the junction and outer bend scour respectively. Downstream migration of the Jamuna–Ganges confluence scour is accompanied by 17 m of deposition at the upstream head of the October 1993 scour (Fig. 1f).

The large fluvial bed scours documented over this 28-month period erode up to 27 m below present sea level. Although subsidence will have allowed accumulation of fluvial sediments here during the period of delta progradation from this region to the present coast^{19,20}, these fluvial scours have incised substantially into the underlying sediments. This may result in the erosive juxtaposition of fluvial sediments on deltaic or shallow-marine deposits²⁰. It is likely that scours of similar magnitude occur nearer the present-day delta mouth in the deltaic distributary channels, again generating superimposition of sediments from markedly different depositional environments. It is clear that substantial scour may thus arise from purely autocyclic processes, in the present case up to about five times the mean channel depth, in agreement with past studies of confluence scour^{18,21,22}. The magnitude of such scours therefore suggests that use of mean channel depth as a criterion for inferring sea-level fall in ancient sediments^{3–10} is inappropriate, and scour depths greater than five times the mean depth must be documented before autocyclic processes can be wholly neglected. This study also demonstrates that these deep river scours are extremely mobile and may migrate across the full width of the braidplain²³ (up to 15 km), thereby controlling the depth and morphology of the basal erosion surface. The erosion surface may also extend beyond the braidbelt width through the influence of channel switching and bed scour by other large rivers that drain the Bengali plain²⁴. Some research²⁵ has suggested a progressive westward shift of the Jamuna over the past 200 years, possibly in response to tectonic tilting, and up to the 1770s the Jamuna drained into the Meghna river some 80 km northeast of the present Jamuna–Ganges junction²⁶. This demonstrates that autocyclic scours, and their erosive bases, may have a far greater regional extent than the braidplain width, again highlighting the need for caution when interpreting such surfaces in the ancient rock record. It should also be noted that soils develop rapidly in abandoned parts of the floodplain and on stable, large (30 × 15 km) braid bar complexes^{23,25}, thus forming 'interfluvial' regions, contemporaneous with the scour surfaces, but away from the true valley margins.

The morphology of the Jamuna–Ganges confluence has a strikingly similar geometry to junctions of smaller (<100 m wide) natural channels^{18,21,22} and those simulated in laboratory experiments^{21,22}, although the bed slopes leading into the Jamuna scours are generally low angle (<5°) without a significant avalanche face. Predictions of scour depth from past studies^{21,22} for the 75° junction angle of the Jamuna–Ganges confluence range between 2 and 4 times the mean channel depth¹⁸. The scale invariance of junction morphology increases the application of small-scale modelling experiments^{21,22} and suggests similarity in the flow and sediment transport processes that control confluence geometry and stability. Because junctions form key nodes within the braided network, such scale invariance may aid our understanding of the fundamental processes that cause channel braiding²⁷.

Recognition of the extent and formative processes of fluvial scour is critical both in devising engineering strategies for large braided rivers^{23,25,28} and in providing reliable and robust process-based criteria for aiding interpretation of ancient sediments and sea-level changes. These data from the Jamuna and Ganges rivers demonstrate that autocyclic scours may fulfil most of the diagnostic criteria used by many workers to infer sequence stratigraphic boundaries and sea-level fall in the ancient rock record. Although

many case studies do convincingly document substantial sea-level fall and the subsequent alluvial response^{3,4,19}, our investigation of large alluvial channels suggests that to define base-level controlled incision unambiguously, several criteria must be examined and satisfied. First, the basal erosion surface must be greater than five times the mean channel depth. Second, an estimate of both channel and braidplain width should be obtained to provide limits for the local extent of basal scour. Third, mean avulsion step length should be estimated to indicate the regional extent of the autocyclic basal erosion surface. Fourth, erosive sequence boundaries should be traceable between basins. Finally, the occurrence of palaeosols, at the same chronostratigraphic level as the erosive surface, may not by itself be diagnostic of a true valley interfluvial and must be examined in relation to the sedimentology of the underlying deposits. □

Received 16 October 1996; accepted 24 March 1997.

1. Postmienter, H. W., Jervey, M. T. & Vail, P. R. Eustatic controls on clastic deposition I—conceptual framework. *Soc. Econ. Paleontol. Mineral. Spec. Publ.* **42**, 109–124 (1988).
2. Postmienter, H. W. & Vail, P. R. Eustatic controls on clastic deposition II—Sequence and systems tract models. *Soc. Econ. Paleontol. Mineral. Spec. Publ.* **42**, 125–154 (1988).
3. Van Wagoner, J. C., Mitchum, R. M., Campion, K. M. & Rahmanian, V. D. Siliciclastic sequence stratigraphy in well logs, cores and outcrops: concepts for high resolution correlation of time and facies. *Am. Assoc. Petrol. Geol. Methods in Exploration* **7**, (1990).
4. Dalrymple, R. W., Boyd, R. & Zaitlin, B. A. (eds) Incised-Valley Systems: Origin and Sedimentary Sequences. *Soc. Econ. Paleontol. Mineral. Spec. Publ.* **51**, (1994).
5. Hampson, G. J., Elliott, T. & Flint, S. S. Critical application of high resolution sequence stratigraphic concepts to the Rough Rock Group (Upper Carboniferous) of northern England. *Geol. Soc. London Spec. Publ.* **104**, 221–246 (1996).
6. Aitken, J. F. & Flint, S. S. The application of high-resolution sequence stratigraphy to fluvial systems: a case study from the Upper Carboniferous Breathitt Group, eastern Kentucky, USA. *Sedimentology* **42**, 3–30 (1995).
7. Flint, S., Aitken, J. & Hampson, G. Application of sequence stratigraphy to coal-bearing coastal plain successions: implications for the UK Coal Measures. *Geol. Soc. London Spec. Publ.* **82**, 1–16 (1995).
8. Davies, S. J. & Elliott, T. Spectral gamma ray characterization of high resolution sequence stratigraphy: examples from Upper Carboniferous fluvio-deltaic systems, County Clare, Ireland. *Geol. Soc. London Spec. Publ.* **104**, 25–35 (1996).
9. Hampson, G. Discrimination of regionally extensive coals in the Upper Carboniferous of the Pennine Basin, UK, using high resolution sequence stratigraphic concepts. *Geol. Soc. London. Spec. Publ.* **82**, 79–97 (1995).
10. Aitken, J. F. & Flint, S. S. Variable expressions of interfluvial sequence boundaries in the Breathitt Group (Pennsylvanian), eastern Kentucky, USA. *Geol. Soc. London Spec. Publ.* **104**, 193–206 (1996).
11. Shanley, K. W. & McCabe, P. J. Perspectives on the sequence stratigraphy of continental strata. *Am. Assoc. Petrol. Geol. Bull.* **78**, 544–568 (1994).
12. Leckie, D., Fox, C. & Tarnocai, C. Multiple paleosols of the late Albian Boulder Creek Formation, British Columbia, Canada. *Sedimentology* **36**, 307–322 (1989).
13. Leeder, M. R. & Stewart, M. D. Fluvial incision and sequence stratigraphy: alluvial responses to relative sea-level fall and their detection in the geological record. *Geol. Soc. London Spec. Publ.* **103**, 25–39 (1996).
14. Leckie, D. A. Canterbury Plains, New Zealand—implications for sequence stratigraphic models. *Am. Assoc. Petrol. Geol. Bull.* **78**, 1240–1256 (1994).
15. Schumm, S. A. & Winkley, B. A. in *The Variability of Large Alluvial Rivers* 1–9 (Am. Soc. Civ. Engrs., New York, 1994).
16. Salter, T. Fluvial scour and incision: models for their influence on the development of realistic reservoir geometries. *Geol. Soc. London Spec. Publ.* **73**, 33–51 (1993).
17. River Survey Project Flood Action Plan 24 Final Report for Flood Plan Coordinating Committee (Delft Hydraulics/Danish Hydraulics Institute/Hydroland/Approtech/Osiris, Dhaka, 1996).
18. Klaassen, G. J. & Vermeer, K. Confluence scour in large braided rivers with fine bed material. *Proc. Int. Conf. Fluvial Hydraulics*, 395–408 (Vituki, Budapest, 1988).
19. Lindsay, J. F., Holliday, D. W. & Hulbert, A. G. Sequence stratigraphy and the evolution of the Ganges–Brahmaputra delta complex. *Am. Assoc. Petrol. Geol. Bull.* **75**, 1233–1254 (1991).
20. Umitzu, M. Late Quaternary sedimentary environments and landforms in the Ganges delta. *Sedim. Geol.* **83**, 177–186 (1993).
21. Ashmore, P. E. & Parker, G. Confluence scour in coarse braided streams. *Wat. Resour. Res.* **19**, 392–402 (1983).
22. Best, J. L. Sediment transport and bed morphology at river channel confluences. *Sedimentology* **35**, 481–498 (1988).
23. Klaassen, G. J., Mosselman, E. & Brühl, H. On the prediction of planform changes in braided sand-bed rivers. *Proc. Int. Conf. Hydrosci. Engng.* 134–146 (Centre for Comp. Hydroscience and Engng, Mississippi, 1993).
24. Barua, D. K. On the environmental controls of Bangladesh river systems. *Asia Pacif. J. Envir. Dev.* **1**, 81–98 (1994).
25. Thorne, C. R., Russell, A. P. G. & Alam, M. K. Planform pattern and channel evolution of the Brahmaputra River, Bangladesh. *Geol. Soc. London Spec. Publ.* **75**, 257–276 (1993).
26. Winkley, B. R., Leslighter, E. J. & Cooney, J. R. in *The Variability of Large Alluvial Rivers* 269–284 (Am. Soc. Civ. Engrs., New York, 1994).
27. Murray, A. B. & Paola, C. A cellular model of braided rivers. *Nature* **371**, 54–57 (1994).
28. Peters, J. J. Morphological studies and data needs. *Proc. Int. Workshop on Morphological Behaviour of Major Rivers in Bangladesh* (Dhaka, 1993).

Acknowledgements. The data used here were collected by the River Survey Project (FAP24), a project conducted by Delft Hydraulics and the Danish Hydraulics Institute as main contractors, the Flood Plan Coordinating Organisation (FPCO) as client, and with the European Union as funding agency. Permission to use the data is gratefully acknowledged. We thank G. J. Klaassen, J. Grijnsen, M. van der Wal, J. J. Peters, H. Hoyer, D. K. Barua, Z. Khan, P. van Groen, C. Iversen and K. Kyhl for advice, co-operation and assistance in the River Survey Project, the School of Geography Graphics Unit for production of the colour figure, and P. Wignall and M. Leeder for comments on this work.

Correspondence and requests for materials should be addressed to J. L. Best (e-mail: J.Best@earth. leeds.ac.uk).

Synaptic plasticity in a cerebellum-like structure depends on temporal order

Curtis C. Bell*, Victor Z. Han*, Yoshiko Sugawara† & Kirsty Grant‡

* R. S. Dow Neurological Sciences Institute, Good Samaritan Hospital and Medical Center, 1120 N.W. 20th Avenue, Portland, Oregon 97209, USA

† Institut Alfred Fessard, CNRS 91190 Gif-sur-Yvette, France

‡ Department of Physiology, Teikyo University School of Medicine, Kaga 2-11-1, Itabashi-ku, Tokyo 173, Japan

Cerebellum-like structures in fish appear to act as adaptive sensory processors, in which learned predictions about sensory input are generated and subtracted from actual sensory input, allowing unpredicted inputs to stand out¹⁻³. Pairing sensory input with centrally originating predictive signals, such as corollary discharge signals linked to motor commands, results in neural responses to the predictive signals alone that are 'negative images' of the previously paired sensory responses. Adding these 'negative images' to actual sensory inputs minimizes the neural response to predictable sensory features. At the cellular level, sensory input is relayed to the basal region of Purkinje-like cells, whereas predictive signals are relayed by parallel fibres to the apical dendrites of the same cells⁴. The generation of negative images could be explained by plasticity at parallel fibre synapses⁵⁻⁷. We show here that such plasticity exists in the electrosensory lobe of mormyrid electric fish and that it has the necessary properties for such a model: it is reversible, anti-hebbian (excitatory postsynaptic potentials (EPSPs) are depressed after pairing with a postsynaptic spike) and tightly dependent on the sequence of pre- and post-synaptic events, with depression occurring only if the postsynaptic spike follows EPSP onset within 60 ms.

Plasticity at parallel fibre synapses was tested in an *in vitro* slice preparation from the electrosensory lobe of the mormyrid electric fish *Gnathonemus petersii*. Afferents from electroreceptors terminate in the electrosensory lobe (Fig. 1a). All recordings were taken from GABAergic Purkinje-like cells^{8,9} (Fig. 1a, inset) which are also known as medium ganglion cells (47 cells; mean membrane potential, 62 mV; range, 53–80 mV; s.e.m., 1.2 mV). Recordings from these cells show large broad spikes (40–60 mV, 8–15 ms; Fig. 1b) that probably arise in the apical dendrites. The spikes are sodium rather than calcium spikes because they are blocked by tetrodotoxin but are still present in a medium with zero calcium and/or cadmium (200 μ M) to block calcium channels (7 cells; Fig. 1b). A previous *in vivo* study indicated that the occurrence of the broad spikes is necessary for a plastic change in the response of these cells to a corollary discharge linked to the motor command that elicits the electric organ discharge⁵.

Transverse slices of ELL were prepared and two stimulating electrodes (S1 and S2) were placed in the molecular layer to activate separate sets of parallel fibres (Fig. 1a). Parallel fibre stimuli evoke an EPSP lasting for 30 to 60 ms (Fig. 1b) or an EPSP followed by an inhibitory postsynaptic potential (IPSP; mean EPSP amplitude before any pairings, 2.8 mV; range, 0.2–6.8 mV; s.e.m., 0.24 mV). S1 and S2 parallel fibre inputs were stimulated with a fixed interval of 500 or 250 ms between them at a rate of 0.1 Hz during control periods before and after pairing, and at a rate of either 1 Hz (42 cells with an S1–S2 interval of 500 ms) or 2 Hz (5 cells with an S1–S2 interval of 250 ms) during pairing. During a pairing, the broad spike was evoked at a fixed moment in the cycle by a 20-ms depolarizing current pulse to the postsynaptic cell. No differences were observed between pairings at 1 and 2 Hz and the data from these two

protocols was pooled. Pairing was maintained for 6 min (360 pairings at 1 Hz) and as many as six different pairings were carried out in the same cell, each pairing with a different timing relation between the intracellularly evoked broad spike and stimulation of the S1 and S2 parallel fibre inputs. Two types of controls were carried out: a 'broad spikes only' control in which the postsynaptic broad spikes were evoked at 1 Hz with no concomitant parallel fibres EPSPs; and an 'EPSP only' control, in which the parallel fibre EPSPs were evoked by S1 and S2 at 1 Hz, with no concomitant postsynaptic broad spike. Parallel fibre EPSP amplitudes were measured for 6–15 min before and after pairings and controls. Differences were tested for significance using the two tailed Student's *t*-test.

Most pairings resulted in significant changes in EPSP amplitude. Forty three of the 47 cells tested had at least one pairing with a significant change in EPSP size. Figure 2A shows a pairing in which the broad spike was evoked 30 ms after the onset of the S1 EPSP and 215 ms before the S2 EPSP. The S1 EPSP decreased after the pairing, whereas the S2 EPSP increased (Fig. 2B).

EPSPs evoked by the same parallel fibre stimulus could increase or decrease in amplitude, depending on the timing relation to the

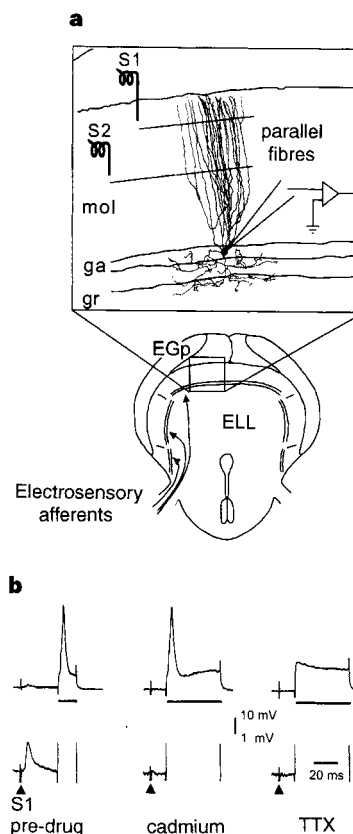


Figure 1 Experimental arrangement and demonstration that the broad spike is a sodium spike. **a**, Transverse section through the electrosensory lobe; inset shows the experimental arrangement. The cell is a reconstruction of a biocytin-filled Purkinje-like neuron. Abbreviations: EGp, eminentia granularis posterior; ELL, electrosensory lobe; ga, gr and mol, ganglionic, granule and molecular layers of the electrosensory lobe, respectively. **b**, Intracellular recording from a Purkinje-like cell showing EPSP evoked by S1 stimulus (arrowhead) and broad spike evoked by intracellular current pulses (bars below upper traces indicate 0.4, 0.4 and 0.8 nA for left, middle and right columns, respectively). Upper traces are low gain, lower traces are high gain. Bath application of 200 μ M cadmium causes the S1-evoked EPSP to disappear within a few minutes but the broad spike is unaffected after 20 min (middle column). The broad spike is blocked within a few minutes by bath application of tetrodotoxin (TTX, 0.5 μ M; right column).

broad spike during pairing. The cell shown in Fig. 3, for example, underwent six different pairings (P1–P6; Fig. 3A). The S1 EPSP decreased following pairings in which the broad spike occurred just after the EPSP onset (P2 and P6), but increased after pairings at all other delays (Fig. 3B). Input specificity was shown by the concurrent pairings with S2. The S2 EPSP decreased only after the one pairing (P3) in which the broad spike occurred just after the EPSP onset, but increased or didn't change after pairings at other delays.

Thus, pairings in which the broad spike was delivered between 0 and 60 ms after EPSP onset resulted in EPSP decreases, whereas pairings at other tested delays usually resulted in EPSP increases. Figure 4a summarizes the data for all 188 pairings in the 43 cells which showed at least one significant effect of pairing. The effect of timing is particularly striking for pairings near zero delay (Fig. 4b). Pairings with the broad spike between 0 and 60 ms after EPSP onset yielded depression, whereas pairings with the broad spike between 8 and 90 ms before EPSP onset yielded enhancement. A change in timing of less than 10 ms could change the effect of pairing from depression to enhancement. Neither depression nor enhancement depended on a prior pairing-induced change in the opposite direction.

EPSPs were often followed by IPSPs and changes in IPSP latency or amplitude could affect the measurement of EPSP size, raising the question as to whether plasticity occurs in EPSPs, IPSPs or both. Figure 3C shows three superimposed responses to S1 recorded before pairing 1 (trace a), after pairing 1 (trace b) and after pairing 2 (trace c). Initial positive slopes of the responses were altered by pairing, indicating that the earliest components are changed (Fig. 3C). These early changes probably reflect changes in EPSPs, because the anatomy suggests that only the EPSP is monosynaptic. Plasticity of the EPSP alone was confirmed by recording from 8 cells in which the GABA_A blocker bicuculline (30 μ M) was added to the ACSF. Only EPSPs were present and no hyperpolarizations were visible under these conditions. All sixteen pairings with the broad spike occurring 10 to 30 ms after EPSP onset resulted in significant ($P = 0.01$) decreases in EPSP amplitude of 10 to 66% (mean, 36%; s.e.m., 5%), whereas of fifteen pairings with the broad spike occurring 15 to 30 ms before EPSP onset, fourteen resulted in significant increases in EPSP amplitude of 12 to 100% (mean 44%; s.e.m., 6%).

The IPSPs evoked by parallel fibre stimulation in normal artificial cerebrospinal fluid also appeared to change following pairing,

Figure 2 Plasticity of parallel fibre-evoked EPSPs following pairing with a broad spike. **A**, a, Control responses before pairing (average of 6 sweeps). b Pairing of EPSPs with broad spike evoked by intracellular current (indicated by a bar); the broad spike (asterisk) is partially obscured by the artefact of the current pulse (note lower gain.). c, Responses after 6 min of pairing (average of 6 sweeps). **B**, Graphs showing amplitudes (1-min averages) of S1 and S2 responses before (a) and after (c) pairing (b).

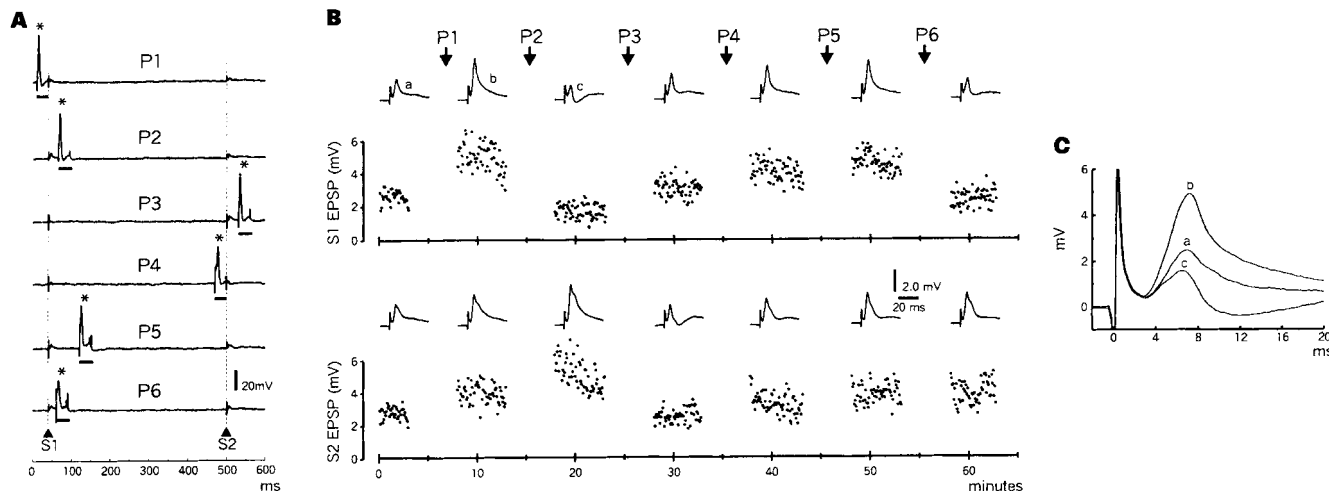
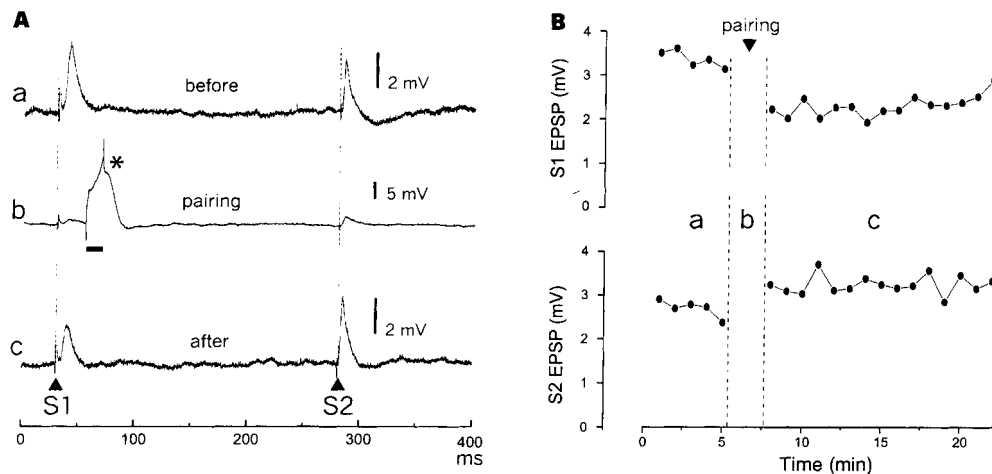


Figure 3 Effects of different pairings in the same cell with different timings of broad spike and EPSPs. **A**, Timing of EPSPs (vertical dashed lines) and broad spike (asterisk). Relative timings of broad spikes with respect to EPSP onset for pairings 1–6 (P1–P6) are as follows (delays with the broad spike after the EPSP are positive and delays with the broad spike before the EPSP are negative): P1 (S1, –29 ms; S2, 490 ms); P2 (S1, 25 ms; S2, –437 ms); P3 (S1, –489 ms; S2, 28 ms); P4 (S1,

429 ms; S2, –32 ms); P5 (S1, 81 ms; S2, 380 ms); P6 (S1, 23 ms; S2, –438 ms). **B**, Graphs showing the peak amplitudes of S1 EPSP (top) and S2 EPSP (bottom) before and after pairings 1–6. Traces above the graphs show the average responses for each period. Traces a, b, and c show responses to S1 before pairing 1, after pairing 1 and after pairing 2, respectively. **C**, Enlarged versions of traces a, b and c are shown superimposed.

increasing when the EPSP decreased and vice versa. Note the appearance of an IPSP to S1 after the second pairing of the cell shown in Fig. 3 (trace c in Fig. 3C) and an IPSP to S2 following pairing 3 (Fig. 3B). Such changes were also seen in cells in which the parallel fibre stimulus evoked a minimal EPSP and a large IPSP. The possibility that an unchanging IPSP is simply unmasked by reduction of the EPSP must still be excluded, however, before plasticity at inhibitory synapses is established.

The EPSP depression observed after pairings in which the broad spike occurred 0 to 60 ms after EPSP onset is clearly an associative effect with a high degree of temporal specificity. The same EPSP was

enhanced after pairing at other delays, and a similar EPSP delivered during the same pairing, but at a delay outside this time window, was also enhanced. The degree to which EPSP enhancement by pairing outside this time window is also associative is less clear. Two controls were carried out to control for non-associative factors. When the EPSP was evoked alone at 1 Hz for 6 min there was clear enhancement, but presentation of the broad spike alone at 1 Hz for 6 min had no effect. The enhancement due to giving the EPSP alone was not significantly different from the enhancement after pairings in which the broad spike was delivered at long delays ('Long-delay pairings' in Fig. 4c), or the enhancement after pairings in which the broad spike was delivered just before the EPSP ('short-delay pairings' (-50 to 0 ms)' in Fig. 4c). (All eight of the other pairwise comparisons among the five control and experimental groups of Fig. 4c showed significant differences at the 0.01 level, after making the Bonferroni adjustment for multiple comparisons.)

The 'EPSP only' control suggests that the enhancements seen with pairings outside the 0–60 ms window might be due to a non-associative effect. A similar non-associative increase in EPSP size following repeated stimulation of parallel fibres alone has been observed in the cerebellum^{10,11}. An associative factor might also contribute to the enhancement observed in the mormyrid electrosensory lobe, however. Such a possibility is suggested by *in vivo* results (C.C.B., A. Caputi and K.G., unpublished observations) and the fact that short-delay pairings (-50 to 0 ms) resulted in significantly larger enhancements than long-delay pairings ($P = 0.01$, as described above). Further experiments will be necessary to distinguish an associative effect.

The sharp temporal constraints on pairings that could lead to depression constitute an important result of this study. Work in the hippocampus^{12–16} and cerebellum^{17–19} has also shown that the magnitude or direction of plastic change at a synapse depends on the temporal relation of pre- and postsynaptic events during pairing. In most of these cases, however, the timing measurements of pre- and postsynaptic events has not been as precise, and temporal constraints have not been as sharp, as in our study. The lack of precise temporal constraints is particularly striking for cerebellar long-term depression, where effective timing relations between presynaptic parallel fibre input and postsynaptic Purkinje cell activation by current injection or climbing fibre input have ranged from presynaptic first by 250 ms (ref. 18) to postsynaptic first by 125 ms (ref. 19). However, dependence on temporal order as tight as in our system has been observed recently in cerebral cortical slices²⁰. In the cerebral cortex, in contrast to our results, pairings with the postsynaptic spike given just after EPSP onset resulted in enhancement, whereas pairings with the postsynaptic just before EPSP onset resulted in depression. The fact that the dependence on temporal order was opposite to that observed in our study probably reflects a different type of plasticity in the cerebral cortex from that in the cerebellum or cerebellum-like structures¹⁷.

Our results and the results from cortical slices both indicate that a simple correlation or covariance rule will not always be a sufficient criterion for the occurrence of hebbian or anti-hebbian plasticity. Precise timing and temporal order of individual pre- and postsynaptic events can be critical. Hebb's original statement of his postulate²¹ implies the importance of temporal order, but this has often been overlooked in more recent work. A modified rule would state that excitatory inputs that arrive just before a postsynaptic spike (and which could therefore have contributed to the occurrence of the spike during normal operation) are enhanced with hebbian plasticity but depressed with anti-hebbian plasticity. The temporal window of ~60 ms within which pairings with the broad spike lead to depression in our study is similar to the duration of parallel fibre-evoked EPSPs. The results are thus consistent with the idea that inputs that could cause or affect the occurrence of a postsynaptic spike may be treated quite differently from inputs that arrive too early or too late to have such an effect.

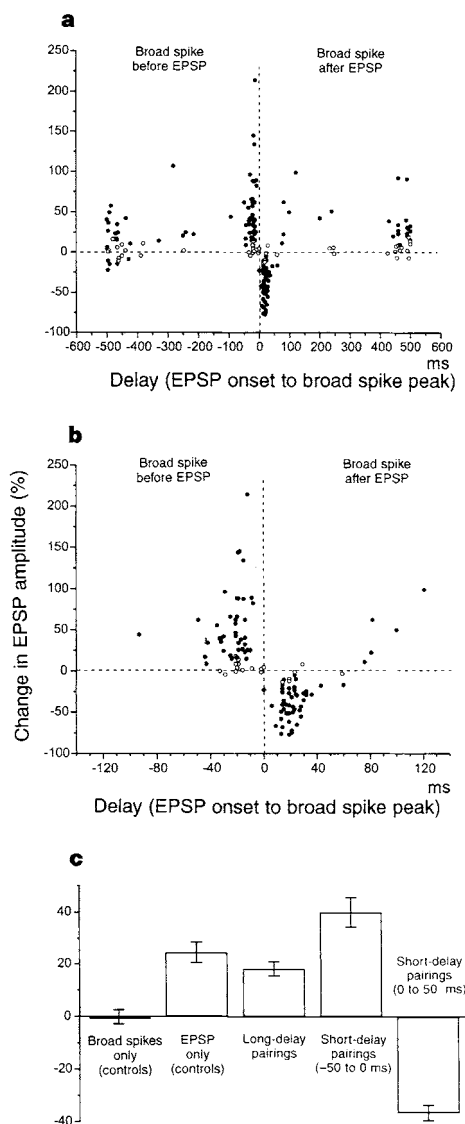


Figure 4 Summary **a**, Percent change in EPSP amplitude plotted against delay between EPSP onset and broad spike peak during pairing. A negative delay with regard to the previous spike and a positive delay with regard to the following spike were present for each pairing. The shorter of these two delays is plotted. Filled circles, significant changes; open circles, non-significant changes (0.01 level). **b**, Same as **a**, but on an expanded scale. **c**, Mean changes in EPSP amplitude following unpaired controls and pairings at different delays (bars show standard error). 'long-delay pairings' includes both positive and negative delays between 300 and 500 ms. The number of controls or pairings was 34, 48, 72, 56 and 51 for the 'broad spike only' control, 'EPSP only' control, long delay pairings, short-delay pairings (-50 to 0 ms), and short-delay pairings (0 to 50 ms), respectively.

The plasticity shown here *in vitro* shares many properties with the plasticity observed *in vivo* following the pairing of a predictive signal with sensory input or intracellular current pulses^{1–6}. In both cases, the changes are input-specific, bidirectional and established within a few minutes. In both cases the plasticity is anti-hebbian, with a decrease in excitation and a possible increase in inhibition taking place within a narrow temporal window. Thus, our results support the hypothesis that anti-hebbian plasticity at the parallel fibre synapse mediates the generation of negative images of predicted sensory input in these structures. □

Methods

400- μ m thick slices of the mormyrid electrosensory lobe were maintained in an interface chamber in humidified 95% O₂, 5% CO₂ and superfused at room temperature (23–25 °C) with artificial cerebrospinal fluid (ACSF), at a rate of 2 ml min⁻¹. The ACSF was saturated with 95% O₂, 5% CO₂ and contained (in mM): NaCl 124; KCl 2.0; KH₂PO₄ 1.25; CaCl₂ 2.6; MgSO₄·7H₂O 1.6; NaHCO₃ 24; glucose 20. Two monopolar insulated tungsten stimulation electrodes were placed in the molecular layer (S1 and S2; inset in Fig. 1a) in order to stimulate two separate parallel fibre inputs to the recorded neurons (pulses of 0.1 ms and 5–50 μ A amplitude). Purkinje-like cells were recorded intracellularly with sharp microelectrodes (150–200 M Ω) containing 2% biocytin dissolved in 2M potassium methyl sulphate. Paired pulse facilitation was measured in 7 cells at interpulse delays of 30–45 ms in order to test for activation of separate groups of fibres by S1 and S2. Facilitation was marked with stimulus pairs S1–S1 (range, 27–108%; mean, 54%) or S2–S2 (range, 18–69%; mean, 42%) but was minimal or absent with stimulus pairs S1–S2 (range, –7 to 13%; mean, –1.3%) or S2–S1 (range, –15 to 3%, mean, –3.6%), indicating that separate groups of fibres were stimulated.

Received 9 September 1996; accepted 24 March 1997.

1. Bell, C. An efference copy which is modified by reafferent input. *Science* **214**, 450–453 (1981).
2. Montgomery, J. & Bodznick, D. An adaptive filter that cancels self-induced noise in the electrosensory and lateral line mechanosensory systems of fish. *Neurosci. Lett.* **174**, 145–148 (1994).
3. Bastian, J. Pyramidal cell plasticity in weakly electric fish: A mechanism for attenuating responses to reafferent electrosensory inputs. *J. Comp. Physiol.* **176**, 63–78 (1995).
4. Montgomery, J., Coombs, S., Conley, R. & Bodznick, D. Hindbrain sensory processing in lateral line, electrosensory, and auditory systems: a comparative overview of anatomical and functional similarities. *Aud. Neurosci.* **1**, 207–231 (1995).
5. Bell, C., Caputi, A., Grant, K. & Serrier, J. Storage of a sensory pattern by anti-Hebbian synaptic plasticity in an electric fish. *Proc. Natl Acad. Sci. USA* **9**, 4560–4564 (1993).
6. Bell, C., Bodznick, D., Montgomery, J. & Bastian, J. The generation and subtraction of sensory expectations within cerebellum-like structures. *Brain, Behav. Evol.* (in the press).
7. Nelson, M. & Paulin, M. Neural simulations of adaptive reafference suppression in the elasmobranch electrosensory system. *J. Comp. Physiol.* **A 177**, 723–736 (1995).
8. Meek, J. et al. Interneurons of the ganglionic layer of the mormyrid electrosensory lateral line lobe: morphology, immunocytochemistry and synaptology. *J. Comp. Neurol.* **375**, 43–65 (1996).
9. Grant, K. et al. Projection neurons of the mormyrid electrosensory lateral line lobe. *J. Comp. Neurol.* **375**, 18–42 (1996).
10. Sakurai, M. Synaptic modification of parallel fiber-Purkinje cell transmission in *in vitro* guinea-pig cerebellar slices. *J. Physiol.* **394**, 463–480 (1987).
11. Hirano, T. Differential pre- and postsynaptic mechanisms for synaptic potentiation and depression between a granule cell and a Purkinje cell in rat cerebellar cultures. *Synapse* **7**, 321–323 (1991).
12. Gustafsson, B., Wigström, H., Abraham, W. & Huang, Y. Long term potentiation in the hippocampus using depolarizing current pulses as the conditioning stimulus to single volley synaptic potentials. *J. Neurosci.* **7**, 774–780 (1987).
13. Debanne, D., Gähwiler, B. & Thompson, S. Asynchronous pre- and postsynaptic activity induces associative long-term depression in area CA1 of the rat hippocampus *in vitro*. *Proc. Natl Acad. Sci. USA* **91**, 1148–1152 (1994).
14. Stanton, P. & Sejnowski, T. Associative long term depression in the hippocampus induced by Hebbian covariance. *Nature* **339**, 215–218 (1989).
15. Hashemzadeh-Gargari, H., Colbert, C. & Levy, W. Refining the temporal defining of an association at the neuronal level using long-term potentiation and long-term depression in the dentate gyrus. *Neurosci. Lett.* **126**, 37–40 (1991).
16. Huerta, P. & Lisman, J. Bidirectional synaptic plasticity induced by a single burst during cholinergic theta oscillation in CA1 *in vitro*. *Neuron* **15**, 1053–1063 (1995).
17. Ito, M. Long term depression. *Annu. Rev. Neurosci.* **85**, 102 (1989).
18. Chen, C. & Thompson, R. Temporal specificity of long-term depression in parallel fiber-Purkinje synapses in rat cerebellar slice. *Learning Mem.* **2**, 185–198 (1995).
19. Ekerot, C. & Kano, M. Stimulation parameters influencing climbing fibre induced long-term depression of parallel fibre synapses. *Neurosci. Res.* **6**, 264–268 (1989).
20. Markram, H., Lübke, J., Frotscher, M. & Sakmann, B. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* **275**, 213–215 (1997).
21. Hebb, D. The organization of behavior. (Wiley, New York, 1949).

Acknowledgements. We thank Y. Fregnac for discussion and Y.F.N. Barmack and R. Kramis for critical reviews of the manuscript. This research was supported by grants from the NSF and the National Institute of Mental Health (to C.B.), by a contract from the European Economic Community, a grant from NATO and funds from the Centre National de Recherche Scientifique (to K.G.), and by grants from the del Duca Foundation and Yamada Science Foundation (to Y.S.).

Correspondence and requests for materials should be addressed to C.C.B. (e-mail: bellc@lhs.org).

Dynamics of orientation tuning in macaque primary visual cortex

Dario L. Ringach, Michael J. Hawken & Robert Shapley

Center for Neural Science, New York University, 4 Washington Place, New York, New York 10003, USA

Orientation tuning of neurons is one of the chief emergent characteristics of the primary visual cortex, V1 (refs 1, 2). Neurons of the lateral geniculate nucleus, which comprise the thalamic input to V1, are not orientation-tuned, but the majority of V1 neurons are quite selective. How orientation tuning arises within V1 is still controversial^{1,3–17}. To study this problem, we measured how the orientation tuning of neurons evolves with time^{18–20} using a new method: reverse correlation in the orientation domain. Orientation tuning develops after a delay of 30–45 milliseconds and persists for 40–85 ms. Neurons in layers 4C α or 4C β , which receive direct input from the thalamus, show a single orientation preference which remains unchanged throughout the response period. In contrast, the preferred orientations of output layer neurons (in layers 2, 3, 4B, 5 or 6) usually change with time, and in many cases the orientation tuning may have more than one peak. This difference in dynamics is accompanied by a change in the sharpness of orientation tuning; cells in the input layers are more broadly tuned than cells in the output layers. Many of these observed properties of output layer neurons cannot be explained by simple feedforward models^{1,3–6}, whereas they arise naturally in feedback networks^{7–17}. Our results indicate that V1 is more than a bank of static oriented filters; the dynamics of output layer cells appear to be shaped by intracortical feedback.

We adapted the reverse correlation method²¹ to study the dynamics of orientation tuning in macaque V1. The stimulus was generated as follows. For each cell, a set *S* of sinusoidal gratings of a fixed spatial frequency but different orientations and spatial phases was generated. In these experiments, the orientation domain was sampled in steps ranging between 3° and 12°. For each orientation, sinusoidal gratings at four spatial phases a quarter of a cycle apart were included in the set. A stochastic image sequence (the stimulus) was generated by randomly selecting, at each video refresh time, a new image from *S*. This was done at a rate of 60 frames per second for a period of 15 min.

The data were analysed to determine the dependence of the cell's response on the past history of oriented gratings presented on its receptive field (Fig. 1). First, an array of counters corresponding to each of the orientations present in the stimulus was zeroed. A fixed value of a time-delay parameter τ was selected. For each nerve impulse recorded, we went back τ ms in time and obtained the orientation of the grating that was present at that moment in the image sequence. The counter corresponding to that orientation was incremented by one. Gratings at the same orientation but different spatial phases shared the same counter. Thus, this procedure averaged across spatial phases. When all action potentials were distributed in the counters, we normalized the resulting histogram by their total number and thereby obtained an estimate of the probability that an orientation θ was present in the stimulus image sequence τ ms before a nerve impulse was generated by the cell. This probability is denoted by $r_\tau(\theta)$. For a fixed value of τ , the function $r_\tau(\theta)$ is a probability distribution on the orientation angles present in the stimulus set. Note that this procedure is equivalent to that of computing the forward correlation: the probability of firing of a cell τ ms after the presentation of a particular orientation in the

stimulus sequence. We studied the dynamics of orientation tuning by looking at how the distributions $r_r(\theta)$ evolve with the time-delay parameter τ .

For $\tau = 0$, for example, we do not expect to see any influence of the input on the output train of action potentials because of the time delay between the retina and the evoked activity of cortical cells, so $r_r(\theta)$ should be a uniform distribution on the orientation domain. Similarly, for very long τ values we also expect to obtain a uniform distribution of $r_r(\theta)$; this is because cells have only finite memory. For intermediate values of τ , a smooth peaked distribution is expected which represents the preference of the cell for particular orientations. There will be one value of τ for which the cell will be better tuned than at any other time; we refer to this value as the 'optimal τ '.

We have applied this method to 61 cells of known laminar location in macaque V1. The results indicate two main patterns of dynamics: 'unimodal' and 'multimodal'. Unimodal dynamics mean that the orientation probability distribution develops from a uniform distribution into a smooth unimodal distribution until some optimal value of τ , and then relaxes back to uniform. Dynamical patterns that do not fit this description are referred to as being multimodal.

Figure 2a, b shows two cells with unimodal dynamics in layers 4C α and 4C β . These distributions are broadly tuned for orientation; even at the optimal τ the probability that a cell responded to a grating at an orientation orthogonal to the preferred one is significantly larger than zero. All 12 cells studied in layers 4C α and 4C β showed an evolution of the orientation probability distribution with unimodal dynamical characteristics like those shown in Fig. 2a, b.

Multimodal dynamics (Fig. 2c–h) were observed in 28 out of 49 cells in supragranular and infragranular layers. The laminar distribution of these cells is summarized in Table 1. Figure 2c illustrates multimodal dynamics in one sharply orientation-tuned neuron from layer 4B. The orientation distribution for this cell at $\tau = 53$ ms before a spike occurred is narrow with a single peak; this peak corresponds to the best orientation as measured with steady state drifting sine gratings. For a time delay of $\tau = 59$ ms, the shape of the probability function is not single-peaked but rather resembles a Mexican hat in the orientation domain: the orientations flanking the optimum orientation became relatively less effective in eliciting a response²². The distribution changes at 71 ms, where the orientation probability function shows a broad peak at an orientation orthogonal to that measured at $\tau = 53$ ms. This means that with a 71-ms delay, the most likely orientation to precede a nerve impulse is orthogonal to the main peak measured at shorter delays. Such dynamical shifts between the preferred and the orthogonal orientation, when going from short to long delay times, were a common finding outside layer 4C. All cells classified as multimodal showed such an 'inversion' of the orientation distribution to some extent. Another example of such an inversion in the orientation distributions of a layer 4B cell is shown in Fig. 2d.

We also observed other patterns of multimodal dynamics, for example, cells whose preferred orientation seemed to drift continuously with time¹⁹. These cells were mainly found in layers 2 + 3 and 5. Figure 2e shows the result obtained from a cell in layer 2 + 3. At a time delay of $\tau = 56$ ms, the best orientation is about 130°, and in the subsequent graphs the peak shifts towards 180° (64 ms), winds around zero degrees (72 ms), and then remains stable at 60° for $\tau \geq 88$ ms. Notice that at $\tau = 88$ ms, the distribution has a mode 90° away from the mode of the distribution at $\tau = 56$ ms. Therefore this cell also exhibits an inversion of the orientation distribution. Similar results from a cell in layer 5 are depicted in Fig. 2f: the preferred orientation of this cell at $\tau = 54$ ms is around 120°, at $\tau = 75$ ms it shifted to 90°, and by $\tau = 82$ ms the peak moved to 55°. Figure 2g shows a very sharply tuned cell in layer 2 + 3. The probability distribution profiles of this cell have a Mexican hat

shape for almost the entire period of the cell's response. The orientations flanking the optimal one have the lowest probability of preceding a spike. At 122 ms, the neuron exhibits an inversion of the distribution. Finally, a particular type of dynamics is observed in some neurons of layer 6 (Fig. 2h): a single-peaked distribution develops until $\tau = 59$ ms, but for time delays larger than 59 ms the distributions become clearly double-peaked. The modes are $\sim 90^\circ$ away, corresponding to orthogonal orientations^{22,23}. The secondary peak even becomes dominant at $\tau = 91$ ms.

The change in the dynamics of orientation tuning, from unimodal to multimodal observed when moving from layer 4C to supragranular and infragranular layers is accompanied by a sharpening of orientation tuning. Cells in layer 4C, which all have unimodal dynamics, are more broadly tuned than cells in other layers^{24,25}. This can already be seen in the examples shown in Fig. 2. We also measured the orientation tuning bandwidth in a relatively large population of $N = 316$ cells (39 located in 4C α , 37 in 4C β , and 240 outside layer 4C) using conventional drifting sine wave gratings. Orientation bandwidth was characterized by the circular variance of the responses²⁵ (see Methods). We found that cells in 4C are significantly less tuned than cells outside layer 4C (Wilcoxon

Table 1 Laminar distribution of cells with 'multimodal' dynamics

Layer	Total cells	'Multimodal'	Percentage
2 + 3	18	7	39
4B	12	11	91
4C	12	0	0
5	9	5	56
6	10	5	50

The total number of cells sampled in each layer and the number and percentage of these cells that exhibited 'multimodal' dynamics are shown. All cells in layer 4C had unimodal dynamics. The largest concentration of multimodal cells is in layer 4B. We have not yet sampled from cells in layer 4A.

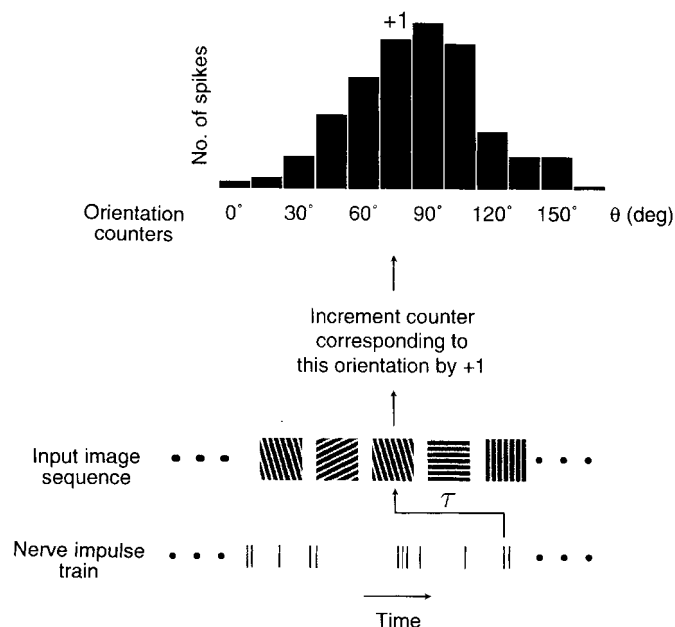


Figure 1 Reverse correlation in the orientation domain. A fast sequence of gratings with a fixed spatial frequency but random orientations and spatial phases is presented in the cell's receptive field. A segment showing five consecutive frames in such a sequence is shown in the figure. The number of times a particular orientation was present in the stimulus sequence τ ms before a spike occurred is shown by the histogram. Normalizing the histogram by the total number of spikes, we obtain an estimate of $r_r(\theta)$, the probability that a grating with orientation θ was present in the stimulus τ ms before an action potential was recorded. We study the dynamics of orientation tuning by looking at the evolution of $r_r(\theta)$ with τ .

rank sum test, $Z = 4.91$, $P < 5 \times 10^{-7}$; the median circular variance in layer 4C was 0.77, and for cells outside 4C was 0.41).

Simulations we have performed of simple feedforward networks based on spatially aligned lateral geniculate nucleus-like inputs (with identical time-course responses for all subunits) show unimodal dynamics (M. Pugh, M. Shelley, D. McLaughlin, R.S. and D.L.R., unpublished observations). The orientation distributions observed outside layer 4C are difficult to reconcile with such a feedforward model. In contrast, some features of the multimodal dynamics arise naturally in recurrent feedback networks^{15–17} (M. Carandini and D.L.R., manuscript in preparation). For instance, inversions of the orientation distributions may occur if neurons with orthogonal orientation preferences inhibit each other²⁶. Centre-surround facilitation and suppression in the orientation domain^{15–17} may account for the secondary peaks seen in distributions like those of the layer 6 cell in Fig. 2h, and the Mexican hat shape of the orientation probability functions in Fig. 2c, g.

The dynamics observed outside layer 4C can be described as having two components: an oriented early excitatory component followed by a delayed inhibitory or suppressive component with similar preferred orientation but broader orientation tuning. This

suppression will sharpen the steady-state orientation tuning of a cell. In some cases (Fig. 2g), the inhibitory component appears quite rapidly and may have an immediate effect on the orientation tuning properties of the cell. It is possible that a more complicated feedforward mechanism than the one we considered is responsible for generating the delayed inhibitory component. However, a simple mechanism like a hyperpolarization of the cell following a recent excitation does not account for most of the dynamical features in the data, such as the Mexican hat distributions or the drifts in the preferred orientation of the cell with time. Long-range lateral interactions between centre and surround could also be contributing to the dynamics or orientation tuning observed in our experiments^{27–29}.

We have observed five cells whose orientation tuning bandwidth decreased with time during the early phase of the response²⁰. In Fig. 3a, normalized distributions from a neuron in layer 2 + 3 for two delay times, $\tau = 45$ ms and 55 ms are shown. There is a clear progression from broad to sharper orientation tuning with time. Figure 3b, c shows further examples from layer 2 + 3 and layer 5. In all cases, sharpening occurred quite rapidly, within 6–10 ms. This phenomenon might be caused in a recurrent network if the

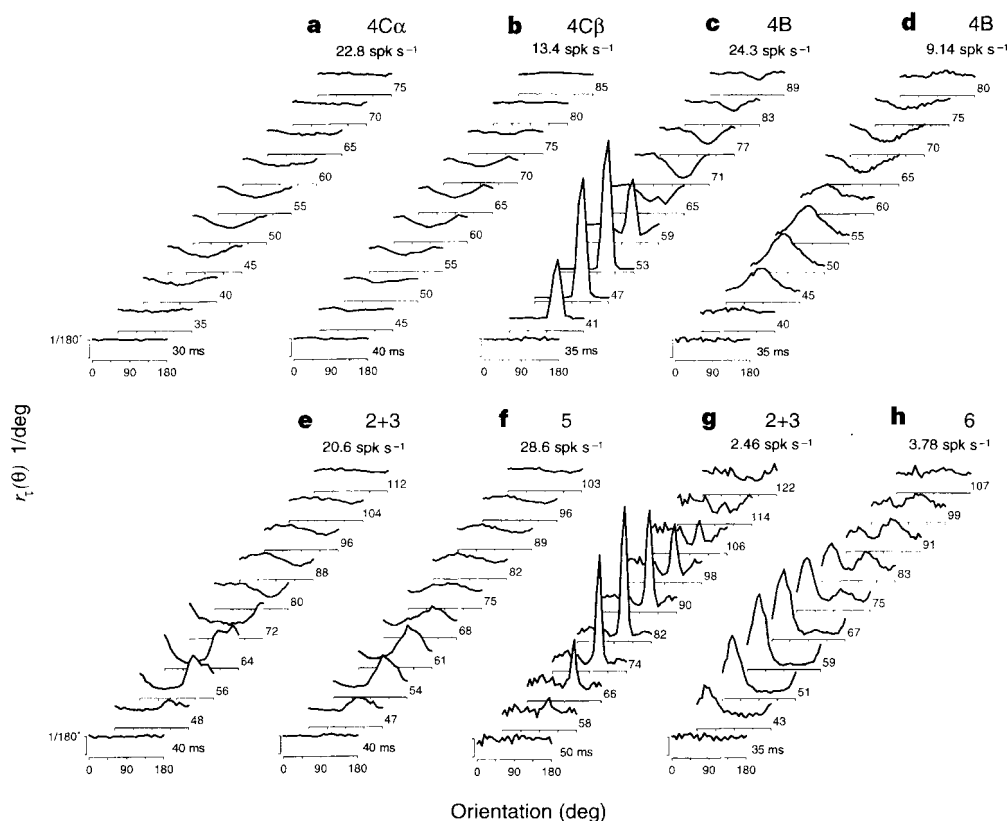


Figure 2 The dynamics of orientation tuning in V1. The graphs within each column show the evolution of the probability distribution function $r_i(\theta)$ with τ . The value of τ is written beside each graph. In all panels the x-axis represents orientation between 0° and 180° . The y-axis represents the value of the probability density function $r_i(\theta)$ for the values of τ indicated. Notice that a uniform distribution has a constant value of $1/180^\circ$ as indicated by the vertical scale at the bottom graph in each column. All graphs have the same vertical scale. As $r_i(\theta)$ is a distribution on the orientation angles for any τ , the integral under all the curves is constant and equal to one. In each case, the distributions $r_i(\theta)$ were uniform up until the first τ shown. The laminar location of each cell and its mean firing rate during the

stimulation period are indicated at the top. **a, b**, Cells in layer 4C α and 4C β show unimodal dynamics and are more broadly tuned than cells in supragranular and infragranular layers. **c**, the result from a cell in layer 4B which is sharply tuned at $\tau = 53$ ms, has a Mexican hat distribution profile at 59 ms, and shows an inversion at 71 ms. **d**, A cell in layer 4B which shows a large inversion of the distribution. **e, f**, Cells in layers 2 + 3 and 5 whose preferred orientation drifted continuously with time. **g**, A sharply tuned cell in layer 2 + 3 whose orientation distribution profile resembles a Mexican hat for most of the response period of the neuron. **h**, Dynamics of a cell in layer 6 which shows the appearance of a secondary peak at the orthogonal orientation for $\tau \geq 67$ ms.

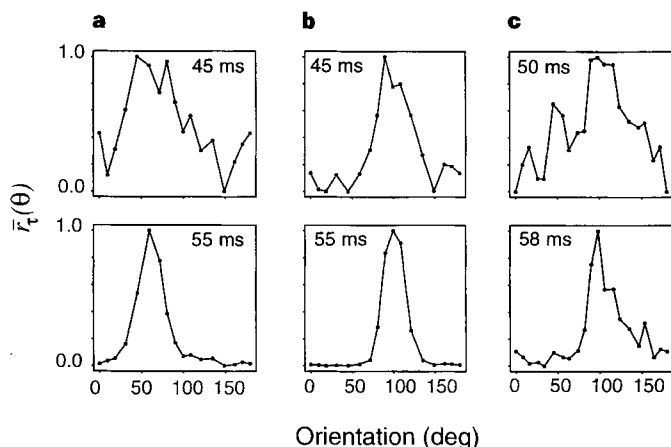


Figure 3 The sharpening of orientation tuning over time. For each given τ , the distribution $r_t(\theta)$ was normalized as follows; $\hat{r}_t(\theta) = \{r_t(\theta) - \min(r_t(\theta))\} / \{\max(r_t(\theta)) - \min(r_t(\theta))\}$; where the minimum and the maximum are taken over the orientation θ . Normalization is required to compare the tuning of the responses independently of their magnitude. Each column in this figure represents the normalized distributions of a cell at two values of τ . **a, b**, Examples of two cells in layer 2 + 3 for which the normalized distributions sharpen over time. **c**, The sharpening effect on a cell in layer 5. This kind of sharpening could be due to recurrent excitatory amplification of the responses at the preferred orientation of the cell¹⁶.

sharpening effect of intracortical feedback is slightly delayed with respect to the input²⁰.

A new and surprising result is that the dynamics of orientation tuning in layer 4C of macaque V1 are quite different from the dynamics in the supragranular and infra-granular layers. The dynamics seen outside layer 4C may be essential in determining the orientation tuning properties of these cells, and could be involved in the encoding of subtler spatial features of the visual image. □

Methods

Acute experiments were performed on adult Old-World monkeys (*Macaca fascicularis*). Animals were initially tranquilized with i.m. acepromazine ($50 \mu\text{g kg}^{-1}$), anaesthetized with i.m. ketamine and maintained on i.v. opioid anaesthetic (sufentanil citrate, $6 \mu\text{g kg}^{-1} \text{ h}^{-1}$). During recording, anaesthesia was continued with sufentanil ($6 \mu\text{g kg}^{-1} \text{ h}^{-1}$) and paralysis induced with pancuronium bromide. Electrocardiogram and expired CO_2 were continuously monitored and blood pressure was measured non-invasively at intervals of 5 min by a Hewlett-Packard Model 78354A patient monitor. Extracellular action potentials were recorded with glass-coated tungsten microelectrodes, exposed tips 5–15 μm . Spikes were detected using a Bak (Maryland, USA) DDIS-1 dual window discriminator and were time-stamped with an accuracy of 1 ms using a CED-1401 Plus (Cambridge, UK) data acquisition system. Strict criteria for single-unit recording included fixed shape of the action potential and the absence of spikes during the absolute refractory period. Small electrolytic lesions ($2\text{--}3 \mu\text{A}$ for 2–3 s, tip negative) were made along the length of each penetration. Details of the reconstruction of the penetrations and the assignment of cells to cortical layers can be found in ref. 30.

A Silicon Graphics Elan R4000 computer generated the stimuli in real time. The screen measured 34.3 cm wide by 27.4 cm high. The refresh rate of the monitor was 60 Hz. The mean luminance of the display was 56 cd m^{-2} . The contrast of the gratings was 100% and their spatial frequency was optimal for each cell. The size of the stimulation patch was large compared to the receptive field of the cell; the side of the stimulus was between 6 and 10 times the spatial period of the optimal grating. The receptive field of the cell was centred in the middle of the stimulus. Therefore, both the classical receptive field of the cell and its surround were stimulated. Most cells responded with mean spike rates

ranging between 2 and 40 spikes per second. A few cells with very high directional selectivity did not respond at all to the stimulus and could not be studied.

We ran stimulus sequences for 15 min (900 s or 54,000 frames). In a typical experiment we used an angular resolution of $\sim 10^\circ$. Thus, the set S usually contained 72 different images ($18 \text{ orientations} \times 4 \text{ spatial phases}$). During the 15-min presentation each image appeared, on average, 750 times. If a typical cortical neuron fired ~ 5 spikes per second to the stochastic stimulation sequence, we obtained a total of 4,500 spikes. We distributed these 4,500 spikes in only 18 orientation bins (because we average across spatial phases). Thus, for a uniform distribution, about 250 spikes are found in each bin. The large number of spikes and small number of orientation bins allowed us to obtain smooth and accurate orientation probability distributions.

The circular variance V of a cell which has responses R_k at angles $0^\circ \leq \theta_k < 180^\circ$ is given by $V = 1 - \frac{|\sum R_k \exp(i2\theta_k)|}{\sum R_k}$. Circular variance is a measure of orientation bandwidth which is bounded between zero and one. Cells not tuned for orientation have a circular variance of one. Cells that are very sharply tuned have circular variance values close to zero.

Received 4 December 1996; accepted 7 March 1997.

- Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture of cat's visual cortex. *J. Physiol. (Lond.)* **160**, 106–154 (1962).
- Hubel, D. H. & Wiesel, T. N. Receptive fields and functional architecture of monkey striate cortex. *J. Physiol. (Lond.)* **195**, 215–245 (1968).
- Chapman, B. et al. Relation of cortical cell orientation selectivity to alignment of receptive fields of the geniculocortical afferents that arborize within a single orientation column in ferret visual cortex. *J. Neurosci.* **11**, 1347–1358 (1991).
- Ferster, D. Orientation selectivity of synaptic potentials in neurons of cat primary visual cortex. *J. Neurosci.* **6**, 1284–1301 (1986).
- Ferster, D. et al. Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature* **380**, 249–252 (1996).
- Reid, R. C. & Alonso, J. M. Specificity of monosynaptic connections from thalamus to visual cortex. *Nature* **378**, 281–284 (1995).
- Andrews, D. P. Perception of contours in the central fovea. *Nature* **205**, 1218–1220 (1965).
- Blakemore, C. et al. Lateral inhibition between orientation detectors in the human visual system. *Nature* **228**, 37–39 (1970).
- Benevento, L. A. et al. Significance of intracortical inhibition in the visual cortex. *Nature* **238**, 124–126 (1972).
- Henry, G. H. et al. Orientation specificity of cells in cat striate cortex. *J. Neurophysiol.* **37**, 1394–1409 (1974).
- Sillito, A. M. et al. A re-evaluation of the mechanisms underlying simple cell orientation selectivity. *Brain Res.* **194**, 517–520 (1980).
- Bonds, A. B. Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex. *Visual Neurosci.* **2**, 41–55 (1989).
- Nelson, S. B. Temporal interactions in the cat visual system. I. Orientation-selective suppression in the visual cortex. *J. Neurosci.* **11**, 344–356 (1991).
- Wörgötter, F. & Koch, C. A detailed model of the primary visual pathway in the cat: comparison of afferent excitatory and intracortical inhibitory connection schemes for orientation selectivity. *J. Neurosci.* **11**, 1959–1979 (1991).
- Somers, D. C., Nelson, S. B. & Sur, M. An emergent model of orientation selectivity in cat visual cortex: simple cells. *J. Neurosci.* **15**, 5448–5465 (1995).
- Douglas, R. J. et al. Recurrent excitation in neocortical circuits. *Science* **269**, 981–985 (1995).
- Ben-Yishai, R. et al. Theory of orientation tuning in the visual cortex. *Proc. Natl Acad. Sci. USA* **92**, 3844–3848 (1995).
- Celebrini, S. et al. Dynamics of orientation coding in area V1 of the awake primate. *Visual Neurosci.* **10**, 811–825 (1993).
- Shevelev, I. A. et al. Dynamics of orientation tuning in the cat striate cortex neurons. *Neuroscience* **56**, 865–876 (1993).
- Pei, X. et al. Receptive field analysis and orientation selectivity of postsynaptic potentials of simple cells in cat visual cortex. *J. Neurosci.* **14**, 7130–7140 (1994).
- de Boer, E. & Kuyper, P. Triggered correlation. *IEEE Trans. Biomed. Enging.* **15**, 169–179 (1968).
- DeValois, R. L. et al. The orientation and direction selectivity of cells in macaque visual cortex. *Vision Res.* **22**, 531–544 (1982).
- Shevelev, I. A. et al. Double orientation tuning in the cat visual cortex units. *Neuroscience* **61**, 965–973 (1994).
- Blasdel, G. G. & Fitzpatrick, D. Physiological organization of layer 4 in macaque striate cortex. *J. Neurosci.* **4**, 880–895 (1984).
- Leventhal, A. G. et al. Concomitant sensitivity to orientation, direction, and color of cells in layers 2, 3, and 4 of monkey striate cortex. *J. Neurosci.* **15**, 1808–1818 (1995).
- Morrone, M. C. et al. Functional implications of cross-orientation inhibition of cortical visual cells. I. Neurophysiological evidence. *Proc. R. Soc. Lond. B* **216**, 335–354 (1982).
- Lund, J. S. Anatomical organization of macaque monkey striate visual cortex. *Annu. Rev. Neurosci.* **11**, 253–288 (1988).
- Blakemore, C. & Tobin, E. A. Lateral inhibition between orientation detectors in the cat's visual cortex. *Exp. Brain Res.* **15**, 439–440 (1972).
- Sillito, A. M. et al. Visual cortical mechanisms detecting focal orientation discontinuities. *Nature* **378**, 492–496 (1995).
- Hawken, M. J. et al. Laminar organization and contrast sensitivity of direction-selective cells in the striate cortex of the Old World monkey. *J. Neurosci.* **8**, 3541–3548 (1988).

Acknowledgements. This work was supported by the NIH (R.S. and M.J.H.) and the Sloan Foundation. We thank M. Carandini, K. Gegenfurtner, J. Krauskopf, M. Landy, F. Mechler, N. Rubin, G. Sapiro and E. Simoncelli for comments on the manuscript. Part of the data obtained using drifting sinewave gratings was collected by A. J. Parker and M. J. Hawken at Oxford University.

Correspondence and requests for materials should be addressed to D.L.R. (e-mail: dario@cns.nyu.edu).

Representations of odours and odour mixtures visualized in the honeybee brain

Jasdan Joerges, Armin Küttner, C. Giovanni Galizia & Randolph Menzel

Institut für Neurobiologie, Freie Universität Berlin, Königin-Luise-Strasse 28-30, 14195 Berlin, Germany

Most animals depend on the identification of odours to locate food or to find mating partners. To accomplish this, the olfactory system must recognize relative concentrations of a large number of substances by analysing complex patterns of chemoreceptor activations^{1,2}, but how these patterns are represented in the brain is not well understood. Previous studies indicated that odours evoke specific patterns of activity in olfactory sensory centres^{3–7} and led to the hypothesis that single glomeruli in the olfactory bulb of mammals respond to particular receptor types^{8–10}. We made optical recordings *in vivo* in the honeybee brain to investigate neuronal population responses to odourants delivered naturally to the animal. We report here that odours evoked specific spatio-temporal excitation patterns in the antennal lobe, the structural and functional analogue of the olfactory bulb¹¹. Specific ensembles of active glomeruli represent odours in a combinatorial manner. A comparison between different individuals shows remarkable similarities for a pheromone component, but not for general flower odours. Mixtures evoked patterns that were combinations of the single odourant responses. These combinations were not fully additive, however, indicating inhibitory effects on single glomeruli. Such interactions could be crucial for the formation of singular codes for complex odour blends.

Both the olfactory bulb and the antennal lobe have a glomerular structure with lateral connections between glomeruli, and process incoming signals from broadly tuned odourant receptors^{12,11–13}. Recordings of glomerular output neurons in insects and vertebrates indicate distributed odour representations^{14–17}. Evidence accumulated from molecular biological studies suggests that receptor neurons expressing the same receptor protein converge onto single glomeruli^{8–10}. To visualize the nature of odour-elicited activity *in vivo*, we made optical recordings of calcium-dependent fluorescence in the honeybee, *Apis mellifera*. The coding of odour mixtures and the similarity between odour representations in different individuals were of special interest.

The antennal lobe is a spherical structure with 156 glomeruli that are innervated by about 60,000 chemoreceptor axons. The glomeruli lie in one layer in the outer rind of the sphere (Fig. 1A, a)¹⁸ and the preparation allowed us to image those lying on the upper half of the antennal lobe. Stimulation with odours led to spatial patterns of increased fluorescence that can be related to the neuroanatomical structure (Fig. 1A, b). The centres of activity consistently corresponded to single glomeruli. Signals reached peak values of 5% change in fluorescence over background fluorescence, allowing for single-sweep measurements. Successive stimulations with the same odour led to reproducible patterns of activity (Fig. 1A, c, d). Correlation analysis of five repeated single trials for the odour citral confirms the high consistency of the responses (average value of correlation matrix: r , 0.85 ± 0.02 s.e.m., $P < 0.001$; time-span between first and last trial: 1 h).

A distributed odour representation implies that single glomeruli can be activated by different odours. Time courses of signals at selected glomeruli are indeed odour-specific (Fig. 1B, a). They can outlast stimulation for several seconds, presumably as a result of calcium dynamics (also, responses of antennal lobe neurons can last

several seconds after odour stimulus offset¹⁹). Intracellular recordings from olfactory interneurons led to the suggestion that odours are not only represented as spatial activity patterns, but as patterns that evolve over time^{17,19,20}. Although our temporal resolution is too low to reveal fast dynamic effects, we could observe slower spatial dynamics that were odour-specific. Changes of spatial activity distribution over time were also reported in the salamander olfactory bulb⁷. An example for the odour hexanol, where different glomeruli are activated in temporal succession ($\Delta t = 500$ ms) is given in Fig. 1B, b.

The patterns were specific for every odour tested, with distributed but overlapping activities (Fig. 2), confirming that single glomeruli may participate in activity maps of different odours. The natural plant extract carnation, which is a complex blend of many substances, did not lead to more extensive activity than the other five pure substances tested here.

To determine whether the odour-specific glomerular activity patterns are conserved between individuals, we compared the signals of different bees (Fig. 3). Citral led to strong medial activation and to similar overall activation patterns ($n = 8$) (Fig. 3a). The minor differences could depend on slight variations in view and preparation. The representation for citral, therefore, seems to be conserved between individuals. For the tested flower odorants,

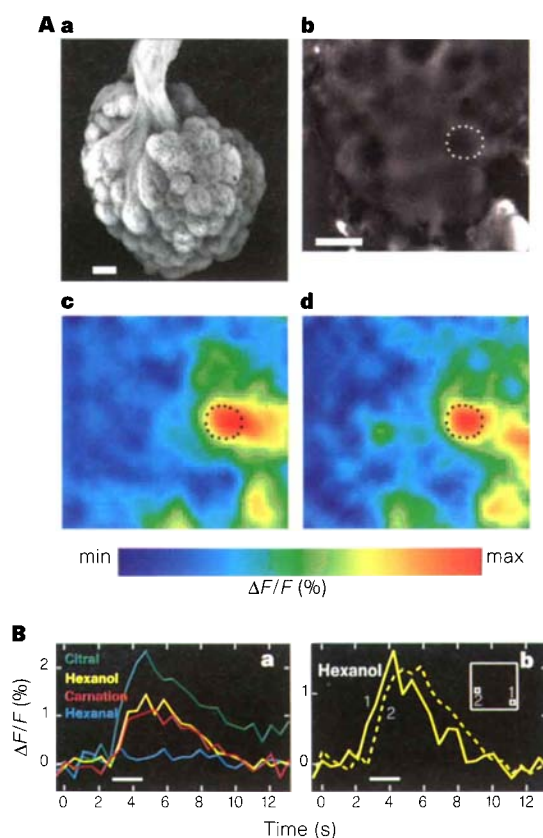


Figure 1 **A**, Optical imaging during stimulation with odours yields spatial activity patterns that are odour specific and can be related to the glomerular morphology of the honeybee antennal lobe (AL). **a**, Reconstruction of an AL. **b**, Fluorescence picture of the imaged area from an AL stained with the dye RH795. A single glomerulus is marked with a dotted circle. Scale bars, 50 μ m. **c**, Odour-evoked activity in the AL shown in **b** during a single 2-s stimulation with citral (1st trial). The outline of the glomerulus marked in **b** is superimposed (dotted line). Signal range $[dF/F]_{min} = -0.1\%$, $[dF/F]_{max} = 1.8\%$. **d**, Citral pattern evoked by stimulation 6 min later (3rd trial with citral, 8th trial in total, $[dF/F]_{min} = 0.1\%$, $[dF/F]_{max} = 1.5\%$). Correlation between patterns in **c** and **d**, $r = 0.92$ ($P < 0.001$). **B**, Time courses of calcium signals at selected glomeruli (different experiment). **a**, Odours elicit specific signals upon stimulation (bar). **b**, Hexanol leads to successive activation of different regions (see inset).

the variability between individuals is higher, as shown for hexanol (Fig. 3b) ($n = 5$). We computed the correlation coefficients for all possible pairs of patterns shown (see Methods). For citral, the average correlation coefficient is $r = 0.50$, whereas for hexanol, $r = 0.35$. The higher correlation coefficient for citral shows that citral induces more constant responses than hexanol. This difference could be related to the fact that citral (besides being a component of some flower fragrances) is an important pheromone component for honeybees. In male moths, specialized macroglomeruli are exclusively concerned with processing sex pheromone signals²¹. Worker honeybees are female and, although pheromones are important for their communication, do not possess a macroglomerular complex. The patterns evoked by citral (and other pheromones, for example isoamyl acetate; see Fig. 2) overlap considerably with those of other tested odours, indicating that certain pheromones are processed by the same neural circuits as general odours.

The olfactory system must provide a concentration-invariant code for odour identity. Bees, for example, recognize odours over a wide range of concentrations²². We investigated the effect of odour intensity by comparing signals evoked at different concentrations. In all cases, patterns evoked by low concentrations were topologically nearly identical to those evoked by high concentrations, but with reduced signal amplitude. For citral, the correlation between high and low concentration responses was $r = 0.90$, $P < 0.001$; for hexanol it was $r = 0.83$, $P < 0.001$. The correlation between citral and hexanol responses at high concentrations was small ($r = 0.29$, $P < 0.01$), and at low concentrations there was no significant correlation between them. Using higher amounts of odour substance ($>10 \mu\text{l}$) led to higher signal amplitudes, showing that our

standard concentration does not lead to saturation. Our findings show that concentration-invariant patterns of activity already occur at the level of the antennal lobe glomeruli, as indicated by results from the salamander olfactory bulb⁷.

Most natural odours are complex blends of volatile compounds, so we studied the patterns evoked by mixtures of odorants. Figure 4a shows the results from stimulation with citral, hexanol and their binary mixture. The mixture pattern included components of both single odours. The reaction profiles shown in Fig. 4b, however, indicate inhibitory interactions at several glomeruli for the mixture response. Mixtures of more than two components led to stronger inhibitory effects. The results in Fig. 4c show that for binary mixtures the inhibitory effects are moderate. For the ternary mixture the inhibitory effects increase, leading to a pattern that differs more from the arithmetic sum of its components than for binary mixtures.

Membrane-permeant and dextran-coupled calcium dyes were used previously in vertebrate preparations^{23,24,31}. We have succeeded for the first time in registering spatio-temporal calcium signals from a central brain structure *in vivo* during natural stimulation. We show that odours are represented in the antennal lobe by specific maps of active glomeruli^{1,3,5}. The acetoxymethyl-ester-coupled dye is likely to be taken up by receptor cells, local interneurons, projection neurons and glial cells. It is important to have some information on the origin of the signals measured. The dense set of afferents from the antenna (Fig. 1A, a) constitutes the largest group of neurons innervating the antennal lobe and these axons exclusively arborize in the rind of glomeruli, whereas other neurons branch in all parts of the glomeruli. This indicates that the afferents

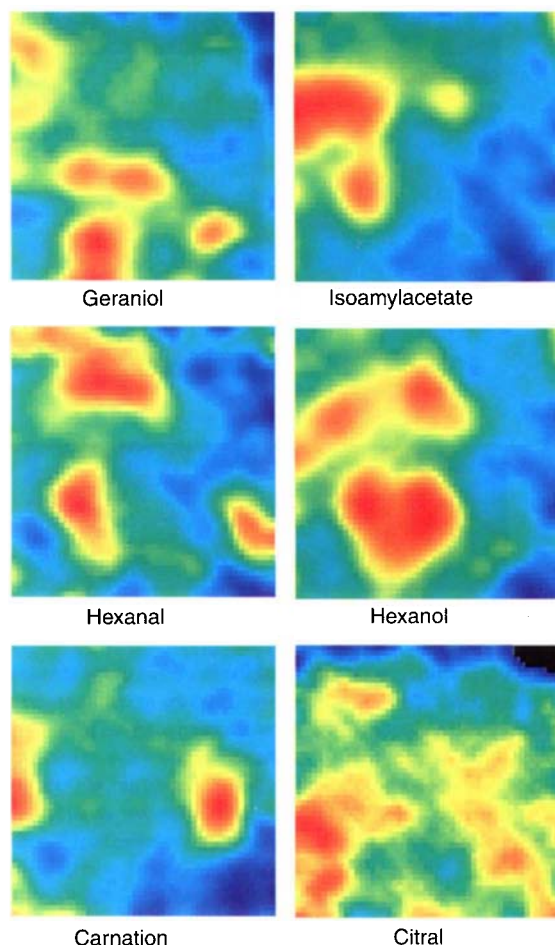


Figure 2 Spatial activity patterns are specific for each tested odour. Six tested odours evoked distributed but overlapping glomerular activity. Neighbouring glomeruli can react quite differently to odours (see central region for hexanol and hexanal, where two glomeruli are active for hexanol, but only one is active for hexanal). Signal ranges (4th trial for each odour): geraniol, $[dF/F]_{\min} = -0.1\%$, $[dF/F]_{\max} = 0.6\%$; isoamylacetate, $[dF/F]_{\min} = -0.6\%$, $[dF/F]_{\max} = 1.9\%$; hexanal, $[dF/F]_{\min} = 0.1\%$, $[dF/F]_{\max} = 1.6\%$; hexanol, $[dF/F]_{\min} = -0.3\%$, $[dF/F]_{\max} = 1.8\%$; carnation, $[dF/F]_{\min} = -0.2\%$, $[dF/F]_{\max} = 1.0\%$; citral, $[dF/F]_{\min} = -0.4\%$, $[dF/F]_{\max} = 0.7\%$.

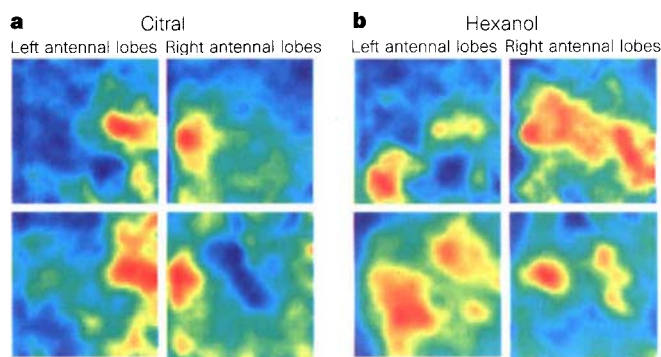


Figure 3 Patterns are more constant in different animals for the pheromone component citral than for the flower odorant hexanol. **a**, Citral patterns from 4 individual honeybees (2 left and 2 right ALs) all showing activity on the central medial side of the AL. Patterns are symmetrical across the longitudinal body axis. Signal ranges: top left, $[dF/F]_{\min} = 0\%$, $[dF/F]_{\max} = 1.1\%$ (single stimulation, 1st trial); top right, $[dF/F]_{\min} = 0.2\%$, $[dF/F]_{\max} = 1.7\%$ (single stimulation, 1st trial); bottom left, $[dF/F]_{\min} = 0\%$, $[dF/F]_{\max} = 0.3\%$ (trials 1–8 averaged); bottom right, $[dF/F]_{\min} = -0.4\%$, $[dF/F]_{\max} = 0.7\%$ (trials 1–4 averaged); **b**, Hexanol patterns from the same 4 animals (same views of the AL) are more variable across individuals (see text). Signal ranges (number of averaged trials as for corresponding citral patterns): top left, $[dF/F]_{\min} = -0.3\%$, $[dF/F]_{\max} = 0.9\%$; top right, $[dF/F]_{\min} = 0.2\%$, $[dF/F]_{\max} = 1.1\%$; bottom left, $[dF/F]_{\min} = 0\%$, $[dF/F]_{\max} = 0.5\%$; bottom right, $[dF/F]_{\min} = -0.2\%$, $[dF/F]_{\max} = 1\%$.

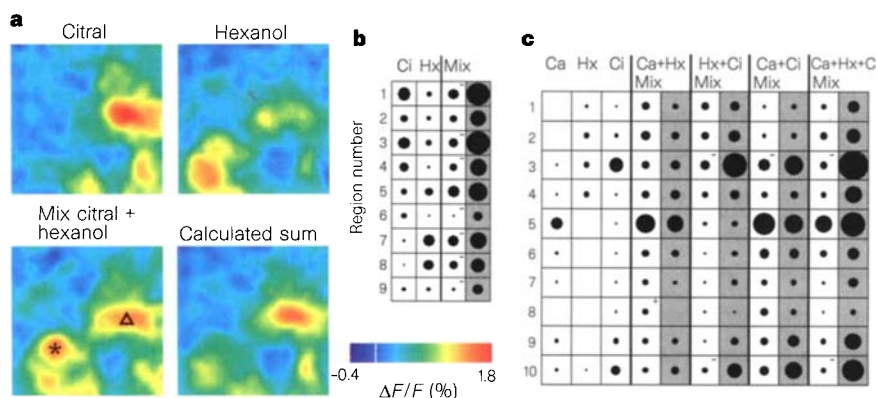


Figure 4 Mixtures of odorants elicited patterns that include components of the single odorant patterns. **a**, Patterns evoked by citral, hexanol and by their binary mixture. For comparison, the arithmetic sum of the two single response patterns is shown. The glomerulus marked with a triangle in the mixture pattern was less active than during stimulation with citral alone. The star marks a glomerulus that was more active than expected by the arithmetic sum (1st trial). **b**, Response profiles of 9 different glomeruli (same experiment as in **a**). Size of circles gives maximal signal in response to odours (smallest circle corresponds to

$dF/F = 0.3\%$; largest circle, $dF/F = 2.1\%$; signals $<0.3\%$ excluded). Activities of citral (Ci) and hexanol (Hx) are combined in the mixture response (Mix). Minus signs mark sites where mixture activity is less than the largest single component activity. The arithmetic sum of Ci and Hx is shown with grey background. **c**, Response profiles from another experiment with carnation (Ca), hexanol and citral. The 3 possible binary mixtures and the ternary mixture are shown as well ($dF/F_{min} = 0.2\%$, $dF/F_{max} = 1.4\%$; trials 1–4 averaged).

contribute a major part to the signal. Mixture interactions between odorants are likely to occur at several stages in the olfactory pathway starting with the receptor cells²⁵. GABA-immunoreactive interneurons of the antennal lobe may also synapse back onto receptor cell terminals, as has been shown in the cockroach *Periplaneta americana*²⁶. Consequently, the observed nonlinear mixture effects could be the result of interactions at the receptor level, or of network interactions in the antennal lobe^{15,27}, or both. In all cases they could be important for the formation of distinct olfactory codes for given odour blends. Even though olfactory receptor cells can have highly nonlinear reaction properties^{25,28}, the patterns were almost independent of stimulus concentration⁷. We conclude that the antennal lobe is capable of forming stable representations from complex sensory responses². Results from molecular biological studies in vertebrates have shown consistent and mirror symmetric projection patterns of olfactory receptors to the olfactory bulb in different individuals^{8,9,10}. The functional activity patterns shown here demonstrate invariance and mirror symmetry for a behaviourally important pheromone. For non-pheromones, the inter-individual differences are higher. A possible hypothesis is that their representations may be modified on the basis of individual experience. Neural plasticity in olfactory neuropiles can lead to changes in behaviour, both in insects²⁹ and vertebrates³⁰. Future experiments, combining functional imaging of the intact brain with olfactory learning, will provide a chance to investigate the formation of olfactory memories *in vivo*. □

Methods

Animals. Adult worker honeybees (departing from the hive) were caught in the morning, fixed in metal tubes and fed with sucrose solution. Animals were experimentally naive, but may have learned odours in the field before being captured.

Preparation, staining and imaging. Calcium-dependent fluorescence was used as an indicator of neuronal activity. Experiments were performed with adult worker bees (in some cases immobilized heads only). The brain was incubated with the calcium indicators calcium-green-1-acetoxymethyl (AM) or calcium-green-2-AM (Molecular Probes, Oregon) for 1 h, resulting in uniform labelling of the antennal lobe (AL). Images of the AL were taken from the intact brain perfused with bee saline through a window cut in the head capsule covered with a cover glass. About 30 of 156 glomeruli in the AL were imaged simultaneously. Clear signals were registered in 27 bees for up to 4 h. Signals are expressed as a change in fluorescence over background fluorescence (dF/F). Series of 30–50 frames (2–3 frames per s) starting 4–6 s before stimulation were recorded with a CCD camera (Photometrics CH250A).

Exposure time per frame was 200 ms. The confocal reconstruction shown in Fig. 1A, a was prepared by injecting neurobiotin into the antennal nerve and staining with streptavidin/CY-3.

Odour stimulation. Odours were delivered to the freely moving antennae using a custom-made olfactometer by switching from a constant stream of air to an odorant containing one in order to eliminate mechanical stimulation associated with stimulus onset. Stimulus duration was 2 s. Concentrations were determined by the amount of odorant substance placed on filter paper inserted in the olfactometer. Pure concentration corresponds to 5 μ l substance, 1:100 corresponds to 5 μ l substance diluted 1:100 in mineral oil. Mixtures were produced in the olfactometer by mixing two odorant-containing air streams (the same results were obtained by using a mixture prepared in advance and placed inside one odour container). To ensure that we used physiological odour concentrations, we trained bees to odours with the same olfactometer (in separate experiments). Animals responded clearly and selectively only to the trained odour (data not shown).

Figures and statistics. Images were spatially low-pass filtered with a 7×7 pixel pyramidal filter. No temporal filters were applied. Data were corrected for dye bleaching by subtracting one sweep without stimulation (background trial) from one with stimulation. For all figures, 3–6 frames during the 2-s stimulation interval were averaged to give a single frame. These were false-colour-coded to show the relative increase of calcium-dependent fluorescence showing the odour-evoked response. For correlation analysis, the linear Pearson correlation coefficient between two frames was calculated on spatially binned data (7×7 pixel). Significances are given for two-sided tests. (Applying Spearman's rank correlation led to equivalent results.) Where measurements of left and right side of the brain were compared (Fig. 3), the mirror image of the left-side patterns and the unmodified right-side patterns were used to calculate correlations to account for the mirror symmetry of left and right. The arithmetic sum pattern shown in Fig. 4a was rescaled to the same range as the measured mixture pattern to allow direct comparison.

Received 18 November 1996; accepted 5 March 1997.

1. Shepherd, G. M. in *Olfaction: A Model System for Computational Neuroscience* (eds Davis, J. & Eichenbaum, H.) 3–41 (MIT Press, Cambridge, MA, 1991).
2. Laurent, G. Odor Images and Tunes. *Neuron* **16**, 473–476 (1996).
3. Stewart, W. B., Kauer, J. S. & Shepherd, G. M. Functional organization of rat olfactory bulb analyzed by the 2-deoxyglucose method. *J. Comp. Neurol.* **185**, 715–734 (1979).
4. Kauer, J. S. Real-time imaging of evoked activity in local circuits of the salamander olfactory bulb. *Nature* **331**, 166–168 (1988).
5. Rodrigues, V. Spatial coding of olfactory information in the antennal lobe of *Drosophila melanogaster*. *Brain Res.* **453**, 299–307 (1988).
6. Lieke, E. Optical recording of neuronal activity in the insect central nervous system: odorant coding by the antennal lobes of honeybees. *Eur. J. Neurosci.* **5**, 49–55 (1993).
7. Cinelli, A. R., Hamilton, K. A. & Kauer, J. S. Salamander olfactory bulb neuronal activity observed by video rate, voltage-sensitive dye imaging. III. Spatial and temporal properties of responses evoked by odorant stimulation. *J. Neurophysiol.* **73**, 2053–2071 (1995).

8. Vassar, R. *et al.* Topographic organization of sensory projections in the olfactory bulb. *Cell* **79**, 981–991 (1994).
9. Ressler, K. J., Sullivan, K. J. & Buck, L. B. Information coding in the olfactory system: evidence for a stereotyped and highly organized epitope map in the olfactory bulb. *Cell* **79**, 1245–1255 (1994).
10. Mombaerts, P. *et al.* Visualizing an olfactory sensory map. *Cell* **87**, 675–686 (1996).
11. Boeckh, J., Distler, P., Ernst, K. D., Hösl, M. & Malun, D. In *Chemosensory Information Processing* (ed. Schild, D.) 201–227 (Springer, Berlin, 1990).
12. Firestein, S., Picco, C. & Menini, A. The relation between stimulus and response in olfactory receptor cells of the tiger salamander. *J. Physiol.* **468**, 1–10 (1993).
13. Schild, D. Signal integration in the olfactory system. *Trends Neurosci.* **17**, 366–367 (1994).
14. Christensen, T. A., Waldrop, B. R., Harrow, I. D. & Hildebrand, J. G. Local interneurons and information processing in the olfactory glomeruli of the moth *Manduca sexta*. *J. Comp. Physiol.* **A 173**, 385–399 (1993).
15. Sun, X., Fonta, C. & Masson, C. Odour quality processing by bee antennal lobe interneurons. *Chem. Senses* **18**, 355–377 (1993).
16. Mori, K. & Yoshihara, Y. Molecular recognition and olfactory processing in the mammalian olfactory system. *Progr. Neurobiol.* **45**, 585–619 (1995).
17. Wehr, M. & Laurent, G. Odour encoding by temporal sequences of firing in oscillating neural assemblies. *Nature* **384**, 162–166 (1996).
18. Flanagan, D. & Mercer, A. R. An atlas and 3-D reconstruction of the antennal lobes in the worker honey bee, *Apis mellifera*. *Int. J. Insect Morphol. Embryol.* **18**, 145–159 (1989).
19. Christensen, T. A., Hildebrand, J. G., Tumlinson, J. H. & Doolittle, R. E. Sex pheromone blend of *Manduca sexta*: responses of central olfactory interneurons to antennal stimulation in male moths. *Arch. Insect Biochem. Physiol.* **10**, 281–291 (1989).
20. Tank, D. W., Gelperin, A. & Kleinfeld, D. Odors, oscillations, and waves: does it all compute? *Science* **265**, 1819–1820 (1994).
21. Hanson, B. S., Ljungberg, H., Hallberg, E. & Löfstedt, C. Functional specialization of olfactory glomeruli in a moth. *Science* **256**, 1313–1315 (1992).
22. Pelz, C., Gerber, B. & Menzel, R. Odorant intensity as a determinant for olfactory conditioning in honeybees: roles in discrimination, overshadowing and memory consolidation. *J. Exp. Biol.* **200**, 837–847 (1997).
23. Yuste, R. & Katz, L. C. Control of postsynaptic Ca^{++} influx in developing neocortex by excitatory and inhibitory neurotransmitters. *Neuron* **6**, 333–344 (1991).
24. O'Donovan, M. J., Ho, S., Sholomenko, G. & Yee, W. Realtime imaging of neurons retrogradely and anterogradely labelled with calcium-sensitive dyes. *J. Neurosci. Meth.* **46**, 91–106 (1993).
25. Akers, R. P. & Getz, W. M. Response of olfactory receptor neurons in honey bees to odorants and their binary mixtures. *J. Comp. Physiol.* **A 173**, 169–185 (1993).
26. Distler, P. GABA-immunohistochemistry as a label for identifying types of local interneurons and their synaptic contacts in the antennal lobes of the american cockroach. *Histochemistry* **93**, 617–626 (1990).
27. Hansson, B. S., Anton, S. & Christensen, T. A. Structure and function of antennal lobe neurons in the male turnip moth, *Agrotis segetum*. *J. Comp. Physiol.* **A 175**, 547–562 (1994).
28. Ache, B. W. Towards a common strategy for transducing olfactory information. *Semin. Cell Biol.* **5**, 55–63 (1994).
29. Hammer, M. & Menzel, R. Learning and memory in the honeybee. *J. Neurosci.* **15**, 1617–1630 (1995).
30. Kendrick, K. M., Levy, F. & Keverne, E. B. Changes in the sensory processing of olfactory signals induced by birth in sheep. *Science* **256**, 833–836 (1992).
31. Friedrich, R. W. & Karsching, S. I. *Neuron* 833–836 (in the press).

Acknowledgements. We thank M. Hammer, M. Gjurfa, R. Yuste and A. Borst for helpful comments on the manuscript and B. Corrette for help with the English. We are grateful to T. Faber for experimental support and to S. McIlwath for providing the preparation of the antennal lobe shown in Fig. 1A,a.

Correspondence and requests for materials should be addressed to J.J. (e-mail: joerges@neuro.biologie.fu-berlin.de).

Presenilin 1 is required for *Notch 1* and *Dll1* expression in the paraxial mesoderm

Philip C. Wong^{†‡}, Hui Zheng[‡], Hua Chen[†], Mark W. Becher[†], Dalip J. S. Sirinathsinghji[§], Myrna E. Trumbauer[‡], Howard Y. Chen[‡], Donald L. Price^{†¶}, Lex H. T. Van der Ploeg[‡] & Sangram S. Sisodia^{†¶}

Departments of [†] Pathology, [‡] Neuroscience and [§] Neurology, and [¶] the Neuropathology Laboratory, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

[‡] Department of Genetics and Molecular Biology, Merck Research Laboratories, Rahway, New Jersey 07065, USA

[§] Merck Sharp and Dohme Research Laboratories, Neuroscience Research Center, Terlings Park, Eastwick Road, Essex CM20 2QR, UK

[¶] These authors contributed equally to this work.

Approximately 10% of cases of Alzheimer's disease are familial and associated with autosomal dominant inheritance of mutations in genes encoding the amyloid precursor protein¹, presenilin 1 (PS1)² and presenilin 2 (PS2)^{3,4}. Mutations in PS1 are linked to about 25% of cases of early-onset familial Alzheimer's disease⁵. PS1, which is endoproteolytically processed *in vivo*⁶, is a multipass

transmembrane protein and is a functional homologue of SEL-12 (ref. 7), a *Caenorhabditis elegans* protein that facilitates signalling mediated by the Notch/LIN-12 family of receptors^{8,9}. To examine potential roles for PS1 in facilitating Notch-mediated signalling during mammalian embryogenesis, we generated mice with targeted disruptions of PS1 alleles (PS1^{-/-} mice). PS1^{-/-} embryos exhibited abnormal patterning of the axial skeleton and spinal ganglia, phenotypes traced to defects in somite segmentation and differentiation. Moreover, expression of mRNA encoding Notch1 and Dll1 (delta-like gene 1)¹⁰, a vertebrate Notch ligand, is markedly reduced in the presomitic mesoderm of PS1^{-/-} embryos compared to controls. Hence, PS1 is required for the spatiotemporal expression of Notch1 and Dll1, which are essential for somite segmentation and maintenance of somite borders^{11–13}.

In the PS1 targeting construct, an ~1.9 kilobase fragment containing the second coding exon (exon 4, amino acids 30–113) and flanking intronic segments of the PS1 gene was replaced by a neomycin-resistance gene (Fig. 1a). The linearized targeting vector was electroporated into AB2.1 embryonic stem (ES) cells and two ES cell lines (from a total of 65 clones) with a single targeted allele were used for the generation of PS1^{-/-} mice. Genotyping of

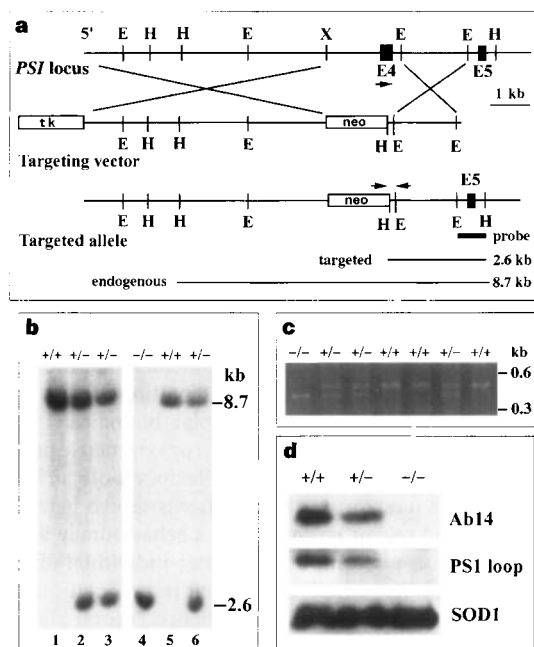


Figure 1 Targeted disruption of the PS1 gene by homologous recombination. **a**, Maps of the wild-type PS1 locus, the targeting vector, and the disrupted PS1 allele. Exons 4 and 5 of PS1 are indicated by black boxes. The targeting vector shows the replacement of exon 4 and flanking genomic sequences by the neomycin gene (neo) and the HSV thymidine kinase gene (tk). Arrows indicate the sites within the targeted and wild-type PS1 alleles from which PCR primers were chosen for genotyping. Lines below denote expected sizes for HindIII-digested fragments detected with a 3'-flanking probe (black bar) from targeted and endogenous PS1 alleles. E, EcoRI; H, HindIII; X, XbaI. **b**, Analysis of genomic DNA from ES cells (1, wild type; 2, clone 300; 3, clone 688; and embryos from PS1^{-/-} crosses (lanes 4–6; genotypes for the PS1 targeted allele are indicated above the lanes). The HindIII fragments detected for wild-type (8.7 kb) and targeted (2.6 kb) PS1 alleles with the 3' probe are indicated. **c**, PCR analysis of DNA extracted from yolk sac. Using primers indicated in **a**, the 370-bp or 500-bp fragment is specific to the targeted or endogenous PS1 allele respectively. **d**, Total protein extracts (100 µg) of wild-type (+/+), heterozygous (+/-) and homozygous PS1 knockout (-/-) E18.5 embryos were immunoblotted using rabbit polyclonal antisera specific for epitopes in the N terminus (Ab14) and the loop region (PS1 loop) of PS1, and superoxide dismutase 1 (SOD1). Bound antibodies were detected with ¹²⁵I-labelled protein A.

$PS1^{-/-}$ mice were performed by DNA blotting (Fig. 1b) and polymerase chain reaction (PCR) methods (Fig. 1c). To confirm that the targeting event led to inactivation of the $PS1$ gene, we prepared total SDS extracts from embryos at embryonic day 18.5 (E18.5) and subjected these preparations to immunoblotting analysis with antibodies specific for epitopes contained within each of the endoproteolytic products of $PS1$ that normally accumulate *in*

*vivo*⁶. In $PS1^{+/-}$ mice, $PS1$ derivatives accumulate to ~50% the level of control littermates, whereas $PS1^{-/-}$ mutant mice showed no evidence of accumulation of $PS1$ -related polypeptides (Fig. 1d). These results confirmed the targeted inactivation of $PS1$.

Although no homozygous mutant mice survived beyond the first day after birth (>50 litters examined), $PS1^{-/-}$ embryos were present at expected mendelian frequencies at various stages of

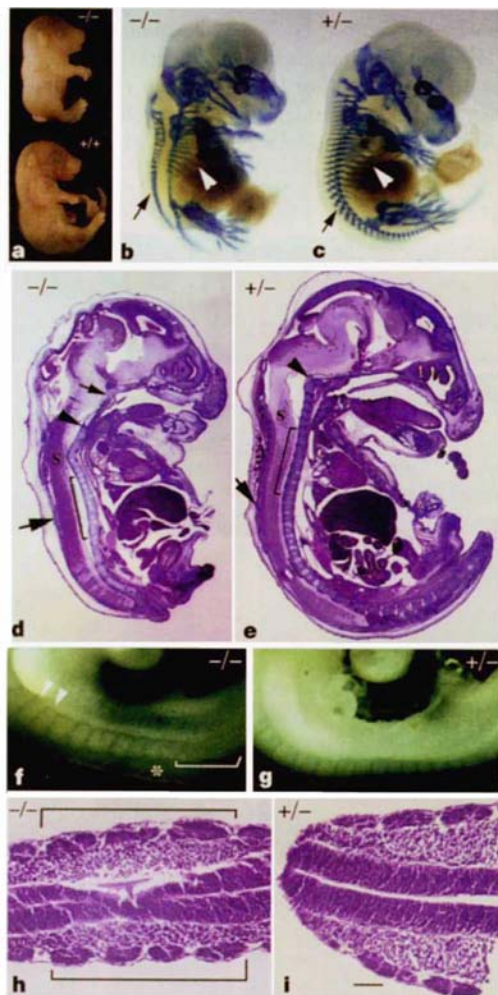


Figure 2 Abnormal patterning of the axial skeleton and somite segmentation defect in $PS1^{-/-}$ embryos. **a**, E17.5 embryos were fixed and photographed intact; note the overall size reduction and the stubby tail of the $PS1^{-/-}$ embryo (top) compared to a littermate control (bottom). **b, c**, Skeletal preparations of alcian blue (cartilage) stained $PS1^{-/-}$ (**b**) and $PS1^{+/+}$ (**c**) E13.5 embryos. Note the vertebral rudiments are fused (arrow) and rib formation (arrowhead) is defective in $PS1^{-/-}$ embryo (**b**), whereas the vertebral column (arrow) and ribs (arrowhead) in littermate control (**c**) are orderly segmented. **d, e**, Sagittal section of E15.5 $PS1^{-/-}$ embryo (**d**) shows abnormal segmentation of vertebral column (bracket) adjacent to spinal cord (s) and fusion of dorsal arches (large arrow) compared to $PS1^{+/+}$ control (**e**). Note the severe bending of the basioccipital bone (small arrow) and the downward disposition of the hindbrain and brainstem in the $PS1^{-/-}$ (**d**) compared to the $PS1^{+/+}$ (**e**) embryo. The arrowhead denotes the distorted angle formed between the basioccipital bone and the atlas in the $PS1^{-/-}$ (**d**) compared to the $PS1^{+/+}$ (**e**) embryo. **f, g**, $PS1^{-/-}$ (**f**) and $PS1^{+/+}$ (**g**) E9.5 embryos were fixed and photographed intact. An ordered array of somites is apparent in $PS1^{+/+}$ embryos (**g**), whereas some somites in $PS1^{-/-}$ embryos (**f**) appear compressed (arrowheads) and fused (asterisks); note the unsegmented condensation of somites (bracketed in **f**). **h, i**, Somite segmentation in $PS1^{+/+}$ (**i**) embryos is coordinated across the midline, whereas asymmetric segmentation of somites is observed in $PS1^{-/-}$ embryo (**h**). Scale bar, 100 μ m.

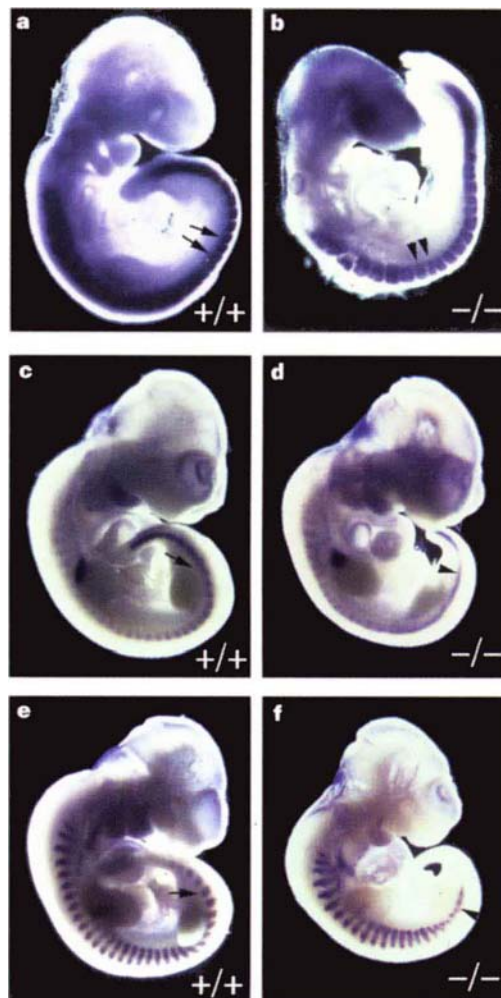


Figure 3 Expression patterns of somitic lineage genes in wild-type and $PS1^{-/-}$ embryos. Detection of *paraxis* (**a, b**), *Pax-1* (**c, d**) and *myogenin* (**e, f**) mRNAs by whole-mount *in situ* hybridization of E9.5 (**a, b**) and E10.5 (**c-f**) embryos. Controls (**a, c** and **e**) and $PS1^{-/-}$ (**b, d** and **f**) embryos are shown. Note the regular staining pattern indicated by arrows in the control embryo in **a**, the irregular pattern of staining (arrowheads) in somites of mutant embryo in **b**, the segmented staining pattern (arrow) in the control embryo in **c**, the unsegmented staining pattern in the caudal region of $PS1^{-/-}$ embryo (arrowhead) in **d**, the metameristic staining pattern denoted by the arrow in the control embryo in **e**, and the unevenly spaced and fused staining pattern in the caudal region of the mutant embryo (arrowhead) in **f**.

gestation from E8.5 to E18.5 (>50 litters examined). We have not yet observed any developmental deficits in mice with a heterozygous mutation of *PS1*. The most striking phenotype observed in *PS1*^{-/-} embryos was a severe perturbation in the development of the axial skeleton. Compared to controls, *PS1*^{-/-} mutant embryos (E10 to E18) are smaller and possess a stubby tail (Fig. 2s). Histochemical analysis of the skeleton of *PS1*^{-/-} E13.5 embryos, using alcian blue stain, revealed considerable defects in the formation of vertebral column and ribs (Fig. 2b,c), although the limbs appeared normal. In mid-sagittal sections from E15.5 *PS1*^{-/-} embryos, we observed that the vertebral column is drastically shortened and fails to undergo proper segmentation (Fig. 2d,e).

The metameric pattern of the axial skeleton, a structure derived entirely from cells in the ventral halves of the somites, is predetermined during somitogenesis^{14,15}. During somite differentiation, cells from the ventral compartment of somites form the mesenchymal sclerotomes, which eventually give rise to vertebral bodies, intervertebral discs, neural arches, pedicles and ribs¹⁴. To determine whether somitogenesis is affected in *PS1*^{-/-} embryos, we examined embryos between E8.5 and E10.5, when somites are being generated. From intact E9.5 *PS1*^{-/-} embryos, we observed irregularly shaped somites along the entire length of the neural tube, although somites were largely absent at the caudalmost regions (Fig. 2f). Further histological examination of E9.5 *PS1*^{-/-} embryos revealed misalignment of somites (Fig. 2h) compared to wild-type embryos in which somites are in tight register across the midline (Fig. 2i). These abnormal somite patterns in *PS1*^{-/-} embryos are highly reminiscent of somite segmentation defects described in mice with functionally inactivated *Notch1* (ref. 11) or *Dll1* (ref. 13) alleles.

To examine somite differentiation in *PS1*^{-/-} embryos, we used whole-mount *in situ* hybridization to analyse the expression of genes that identify specific somitic lineages. *Paraxis*, a gene encoding a basic helix-loop-helix (bHLH) transcription factor, is normally expressed highly in paraxial mesoderm and in newly formed somites¹⁶. In contrast to wild-type E9.5 embryos, in which *Paraxis* is expressed in the entire rostrocaudal array of somites (Fig. 3a), the staining pattern in somites of *PS1*^{-/-} embryos is highly disorganized and irregular (Fig. 3b). *Pax-1*, a gene of the paired box transcription factor family, is specifically expressed in sclerotomal cells in the ventral portion of the somites^{17,18}. In E10.5 *PS1*^{+/+} embryos, *Pax-1* is expressed in a segmented array throughout the ventral sclerotome (Fig. 3c). Although the rostral ventral sclerotome of *PS1*^{-/-} embryos exhibit strong *Pax-1* expression, the staining becomes continuous without defined boundaries in the caudal region beginning at the hind limb-bud level (Fig. 3d). *Myogenin*, a bHLH transcription-factor gene specific to the myogenic lineage, is normally expressed in a metameric pattern along the cranio-caudal axis in the myotome¹⁹. Although the characteristic metameric staining pattern of *myogenin* in the rostral portion of E10.5 *PS1*^{-/-} embryos is similar to that of control embryos (Fig. 3e,f), *myogenin* expression is considerably reduced and the boundaries are not as sharply defined in the caudal region of the mutant embryo (Fig. 3f). These analyses show that, despite the clear disruption in somite segmentation in *PS1*^{-/-} embryos, specification of somitic cell lineages, particularly the sclerotome and dermomyotome, is apparently unaffected.

In view of the similarities in somite segmentation defects in *PS1*^{-/-} embryos and embryos with functionally inactivated *Notch1* (ref. 11) or *Dll1* (ref. 13) alleles, we examined the expression of *Notch1* and *Dll1* mRNA in *PS1*^{-/-} embryos. At E8.5 and E9.5, the abundant expression of *Notch1* mRNA observed in the presomitic mesoderm of control embryos (Fig. 4a,c) is nearly abolished in the *PS1*^{-/-} embryos (Fig. 4b,d). Analysis of *Dll1* mRNA expression show that, although high levels of *Dll1* mRNAs are observed in the presomitic mesoderm of E9.5 control embryos (Fig. 4e), the levels are markedly reduced in *PS1*^{-/-} embryos (Fig. 4f). We confirmed that *PS1* mRNAs are also expressed in the presomitic mesoderm and

somites in wild-type embryos (Fig. 4g), albeit at significantly lower levels than *Notch1* and *Dll1*. These data suggest that *PS1* regulates the spatiotemporal expression of *Notch1* and *Dll1* in the paraxial mesoderm. However, the mechanisms by which *PS1* influences extrinsic or intrinsic signalling pathways necessary for cell-autonomous amplification of *Notch1* or *Dll1* mRNA in the presomitic mesoderm^{8,11} remain to be established.

Because *Dll1* is required for the maintenance of somite borders, such that segment polarity is established in each somite¹³, we examined whether somite polarity is maintained in *PS1*^{-/-} embryos. Histological analysis of *PS1*^{-/-} E11.5 embryos revealed that sclerotome condensation failed to occur (Fig. 4h), suggesting that the identity of the caudal halves of each segment were not specified. As a test of this model we examined the morphogenesis of

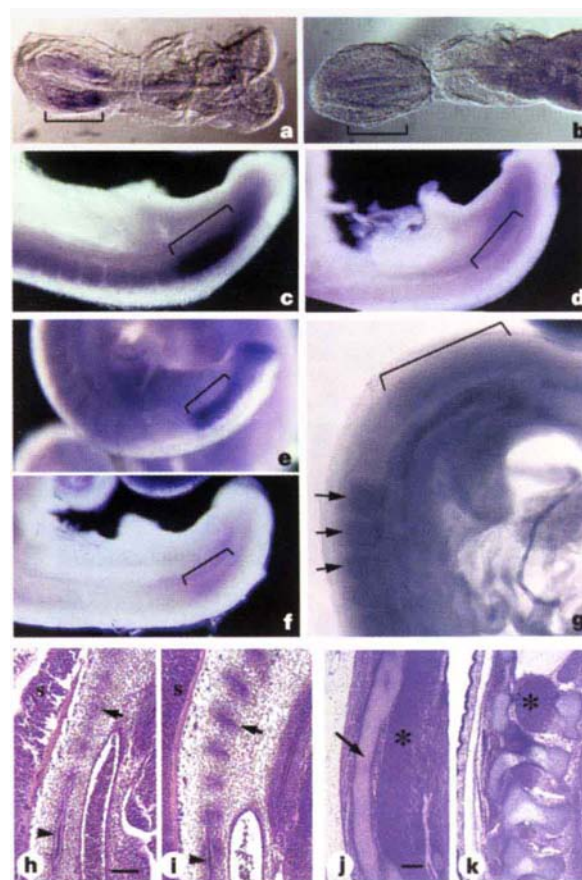


Figure 4 Reduced expression of *Notch1* and *Dll1*, abnormal sclerotome differentiation, and spinal ganglia patterning defects in *PS1*^{-/-} embryos. **a-d**, Detection of *Notch1* mRNA by whole-mount *in situ* hybridization of E8.5 (**a, b**) and E9.5 (**c, d**) embryos. Note the reduction of *Notch1* signals in *PS1*^{-/-} embryos in the presomitic mesoderm (indicated by brackets). **e, f**, Detection of *Dll1* mRNA by whole-mount *in situ* hybridization of E9.5 embryos. Note the decreased *Dll1* signal in *PS1*^{-/-} embryo in the presomitic mesoderm (indicated by bracket). **g**, Detection of *PS1* mRNA by whole-mount *in situ* hybridization of E9.5 embryos. Arrows point to the somites and bracket denotes the presomitic mesoderm. Control (**a, c, e, g**) and *PS1*^{-/-} (**b, d, f**) mice are shown. **h, i**, Sclerotome of *PS1*^{-/-} (**h**) and *PS1*^{+/+} (**i**) embryos at E11.5. The arrow in **b** points to the condensation of sclerotomic material in the *PS1*^{+/+} embryo, whereas the intrasegmental condensation (arrow in **a**) has not occurred in the *PS1*^{-/-} embryo; arrowhead, notochord; s, spinal cord. **j, k**, Parasagittal section of E15.5 *PS1*^{-/-} embryo (**j**) showing fusion of DRG, denoted by an asterisk; the arrow points to the non-segmented axial skeleton. DRG, indicated by an asterisk, are segmented in E15.5 *PS1*^{+/+} embryo (**k**). Sections in **h-k** were stained with haematoxylin and eosin. Scale bars: **h, i**, 200 μ m; **j, k**, 100 μ m.

the dorsal root ganglia (DRG). Classical somite transplantation experiments in chick-quail chimaeric embryos^{20,21} demonstrated that morphogenesis of DRG is intimately governed by the cranio-caudal differentiation of individual somites. In *PS1*^{-/-} embryos, the DRG were fused over multiple segments along the craniocaudal axis of the vertebral column (Fig. 4j,k), strongly suggesting that somite segment polarity is disrupted. Moreover, recent studies¹³ have demonstrated that *Dll1*-null embryos also show abnormal condensation of sclerotome and fused DRG. These data, taken together with our observation that *Dll1* mRNA is severely reduced in the presomitic mesoderm of *PS1*^{-/-} embryos (Fig. 4f), lead us to conclude that PS1 is essential for the function of *Dll1* in establishing somite borders and segment polarity.

Although the cellularity and cytoarchitecture of the developing brain of *PS1*^{-/-} embryos appeared normal, all *PS1*^{-/-} embryos after E11.5 exhibited haemorrhages that were limited to the brain (Fig. 5a) and/or spinal cord (data not shown); these lesions were not seen in *PS1*^{+/-} or *PS1*^{+/+} littermates. The severity of the haemorrhages varied between individual embryos, and seemed to correlate with morbidity (assessed by lack of heartbeat). The haemorrhages are present beneath the primordial dura and leptomeninges, within the ventricles, and in neural parenchyma, rarely with focal necrosis (Fig. 5b). Taken together with the demonstration that haemorrhages in the central nervous system (CNS) are consistently observed in mice with functionally ablated *Dll1* (ref. 13) and *Jagged* (T. Gridley, personal communication) genes, these studies offer support for the view that the encoded molecules facilitate Notch signalling pathways in the development of the CNS vasculature.

Despite our demonstration that PS1 is essential for mouse embryonic development, the function(s) of PS1 during maturation and ageing is not clear, although PS1 is also expressed throughout the adult life of rodents and humans^{2,22,23}. However, autosomal dominant forms of early-onset familial Alzheimer's disease are linked to mutations in PS1 and, by mechanisms presently unclear,

mutant PS1 enhances the production of highly amyloidogenic A β 42 peptides both *in vitro* and *in vivo*²⁴⁻²⁶. These studies suggest that one mechanism by which mutant PS1 predisposes individuals to Alzheimer's disease is by the gain of a deleterious function. Together with the identification of missense mutations in *Notch3* in pedigrees with CADASIL, a disorder characterized by stroke and dementia²⁷ these results suggest that, although perturbations in Notch signalling pathways in embryos lacking genes encoding PS1, Notch or ligands of Notch lead to developmental loss-of-function disorders, mutations in these genes lead to autosomal dominant disorders associated with dementia in adult life. □

Methods

Gene targeting of PS1 gene. Genomic clones containing exon 4 of mouse *PS1* were isolated from a 129/Sv strain of mouse (Lambda FIX II Library, Stratagene, La Jolla, CA). We replaced a 1.9-kb *XhoI/EcoRI* fragment containing exon 4 with the neomycin phosphotransferase gene under the control of the PGK promoter. Introduction of a negative selection marker, the herpes simplex virus thymidine kinase gene, at the 5' end of the construct allowed the use of the positive and negative selection scheme²⁸. The targeting vector was linearized at a unique *BamHI* site before transfection into AB2.1 ES cells, which were subject to double selection, as previously described²⁹. Clones were picked and expanded, and DNA was isolated from a portion of the cells and screened by Southern blot analysis. Frozen cells were expanded and injected into C57B1/6J blastocysts.

Histology and *in situ* RNA analysis. Embryos were fixed in 4% paraformaldehyde for 2 h at 27 °C, dehydrated, embedded in paraffin and sectioned at 10 μ m. Sections were stained with either haematoxylin and eosin or Masson trichrome for histological analysis. Cartilages of embryos were stained with alcian blue³⁰. For RNA detection, embryos were first genotyped by allele specific PCR of yolk-sac DNA and then subjected to whole-mount *in situ* hybridization³⁰.

Received 1 April; accepted 21 April 1997.

1. Busfield, F. & Goate, A. M. in *Pathology of Alzheimer's Disease*, 59–75 (Academic, San Diego, 1995).
2. Sherrington, R. *et al.* Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* **375**, 754–760 (1995).
3. Levy-Lahad, E. *et al.* Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science* **269**, 973–977 (1995).
4. Rogeev, E. I. *et al.* Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene. *Nature* **376**, 775–778 (1995).
5. Cruts, M., Hendriks, L. & Van Broeckhoven, C. The presenilin genes: a new gene family involved in Alzheimer's disease pathology. *Hum. Mol. Genet.* **5**, 1449–1455 (1996).
6. Thinakaran, G. *et al.* Endoproteolysis of presenilin 1 and accumulation of processed derivatives *in vivo*. *Neuron* **17**, 181–190 (1996).
7. Levitan, D. *et al.* Assessment of normal and mutant human presenilin function in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **93**, 14940–14944 (1996).
8. Artavanis-Tsakonas, S., Matsuno, K. & Fortini, M. E. Notch Signaling. *Science* **268**, 225–232 (1995).
9. Levitan, D. & Greenwald, I. Facilitation of *lin-12*-mediated signalling by *sel-12*, a *Caenorhabditis elegans* S182 Alzheimer's disease gene. *Nature* **377**, 351–354 (1995).
10. Bettenhausen, B. *et al.* Transient and restricted expression during mouse embryogenesis of *Dll1*, a murine gene closely related to *Drosophila Delta*. *Development* **121**, 2407–2418 (1995).
11. Conlon, R. A. *et al.* *Notch1* is required for the coordinate segmentation of somites. *Development* **121**, 1533–1545 (1995).
12. Swiatek, P. J. *et al.* *Notch1* is essential for postimplantation development in mice. *Genes Dev.* **8**, 707–719 (1994).
13. Hrabec de Angelis, M., McJintyre II, J. & Gossler, A. Maintenance of somite borders in mice requires the *Delta* homologue *Dll1*. *Nature* **386**, 717–721 (1997).
14. Verbout, A. J. The development of the vertebral column. *Adv. Anat. Embryol. Cell Biol.* **90**, 1–122 (1985).
15. Christ, B. & Wiltling, J. From somites to vertebral column. *Anat. Rev.* **174**, 23–32 (1992).
16. Burgess, R., Cserjesi, P., Ligon, K. L. & Olson, E. N. Paraxis: a basic helix-loop-helix protein expressed in paraxial mesoderm and developing somites. *Dev. Biol.* **168**, 296–306 (1995).
17. Deutsch, U., Dressler, G. R. & Gruss, P. Pax-1, a member of a paired box homologous murine gene family, is expressed in segmented structures during development. *Cell* **53**, 617–625 (1988).
18. Koseki, H. *et al.* A role for Pax-1 as a mediator of notochordal signals during the dorsoventral specification of vertebrae. *Development* **119**, 649–660 (1993).
19. Montarras, D. *et al.* Developmental patterns in the expression of *Myf5*, *MyoD*, *myogenin*, and *MRF4* during myogenesis. *New Biol.* **3**, 592–600 (1991).
20. Stern, C. D. & Keynes, R. J. Interactions between somite cells: the formation and maintenance of segment boundaries in the chick embryo. *Development* **99**, 261–272 (1987).
21. Kalcheim, C. & Teillet, M.-A. Consequences of somite manipulation on the pattern of dorsal root ganglion development. *Development* **106**, 85–93 (1989).
22. Lee, M. K. *et al.* Expression of presenilin 1 and 2 (PS1 and PS2) in human and murine tissues. *J. Neurosci.* **16**, 7513–7525 (1996).
23. Kovacs, D. M. *et al.* Alzheimer-associated presenilin 1 and 2: neuronal expression in brain and localization to intracellular membranes in mammalian cells. *Nature Med.* **2**, 224–229 (1996).
24. Scheuner, D. *et al.* Secreted amyloid β -protein similar to that in the senile plaques of Alzheimer's disease is increased *in vivo* by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. *Nature Med.* **2**, 864–870 (1996).

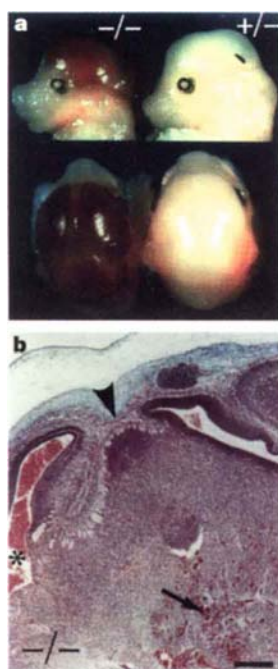


Figure 5 Central nervous system haemorrhage. **a**, A marked brain haemorrhage is evident in the *PS1*^{-/-} (top left (side view) and bottom left (top view)) compared to the *PS1*^{+/-} (top right (side view) and bottom right (top view)) intact E15.5 embryo. **b**, Haemorrhage occurs within ventricles (asterisk), connective tissue overlying the brain (primitive leptomeninges; arrowhead) and parenchyma (arrow). Scale bar, 200 μ m.

25. Duff, K. *et al.* Increased amyloid- β 42(43) in brain of mice expressing mutant presenilin 1. *Nature* **383**, 710–713 (1996).
26. Borchelt, D. R. *et al.* Familial Alzheimer's disease-linked presenilin 1 variants elevate A β 1–42/1–40 ratio *in vitro* and *in vivo*. *Neuron* **17**, 1005–1013 (1996).
27. Joutel, A. *et al.* *Notch3* mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. *Nature* **383**, 707–710 (1996).
28. Mansour, S. L. *et al.* Disruption of the protooncogene *int-2* in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* **336**, 348–352 (1988).
29. Zheng, H. *et al.* β -amyloid precursor protein-deficient mice show reactive gliosis and decreased locomotor activity. *Cell* **81**, 571–480 (1995).
30. Hogan, B., Beddington, R., Costantini, F. & Lacy, E. *Manipulating the Mouse Embryo: A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, NY, 1994).

Acknowledgements. We thank G. Thinakaran, M. Seeger and S. Gandy for PS1 antibodies, C.-M. Fan, A. Gossler, R. Burgess, R. Balling and J. Nye for probes; M. Jiang, S. Gopal and S. Seeburger for their help with tissue culture and blastocyst injection; J. Davis for advice on embryo dissections; V. Gonzalez, M. Peper, G. Crisostomo and J. Van Lare for technical support of histological methods; and A. Gossler, T. Gridley, C.-M. Fan, P. Beachy and J. Gearhart for discussions. This work was supported by the US Public Health Service, the NIH, and by grants from the Adler Foundation (S.S.), the Develbiss Fund (D.L.P. and S.S.) and the Alzheimer's Association (S.S.). D.L.P. is the recipient of a Javits neuroscience investigator award; D.L.P. is the recipient of a Leadership & Excellence in Alzheimer's disease (LEAD) award; S.S. is the recipient of an Alzheimer's Association Zenith award.

Correspondence and requests for materials should be addressed to P.C.W. (e-mail: pcwong@welchlink.welch.jhu.edu).

The role of RhoA in tissue polarity and Frizzled signalling

David I. Strutt, Ursula Weber & Marek Mlodzik

Developmental Biology Programme, EMBL, Meyerhofstrasse 1,
D-69117 Heidelberg, Germany

The tissue polarity genes of *Drosophila* are required for correct establishment of planar polarity in epidermal structures^{1,2}, which in the eye is shown in the mirror-image symmetric arrangement of ommatidia relative to the dorsoventral midline. Mutations in the genes *frizzled* (*fz*), *dishevelled* (*dsh*) and *prickle-spiny-legs* (*pk-sple*) result in the loss of this mirror-image symmetry^{3–5}. *fz* encodes a serpentine receptor-like transmembrane protein required for reception and transmission of a polarity signal^{5,6}. Little else is known of the signalling pathway(s) involved other than that Dsh acts downstream of Fz⁷. We have identified mutations in the *Drosophila* homologue of RhoA p21 GTPase, and by analysis of their phenotype show that RhoA is required for the generation of tissue polarity. Genetic interactions indicate a role for RhoA in signalling mediated by Fz and Dsh, and furthermore suggest that JNK/SAPK-like kinases are involved. These data are consistent with a Fz/RhoA signalling cascade analogous to the yeast pheromone signalling pathway⁸ and that proposed for activation of the serum response factor (SRF) in vertebrate cells⁹.

Rho family p21 GTPases have previously been widely studied as regulators of the actin cytoskeleton¹⁰, and more recently as components of signal transduction cascades¹¹. Studies of Rho family p21 GTPases in *Drosophila* using dominant-negative or constitutively activated forms have suggested possible functions in a variety of developmental contexts including morphological movements in the embryonic epidermis, axon outgrowth and planar polarization in the wing^{12–15}.

We isolated several loss-of-function alleles of *RhoA* (Fig. 1a). Embryos with a homozygous mutation for the null alleles of the *RhoA* gene exhibit an 'anterior open' phenotype, consistent with a defect in morphological movements or cell polarity in the embryonic epidermis (Fig. 1b, c). Homozygous mutant clones of cells carrying null alleles in imaginal tissue fail to proliferate, indicating a requirement for *RhoA* in cell growth and/or viability. However, clones of hypomorphic alleles give rise to mutant tissue with specific patterning defects. In the eye, ommatidia are incorrectly rotated relative to the equator, and sometimes exhibit inappropriate chiral forms (Fig. 2a). This phenotype is very similar to that of the tissue

polarity mutants. Loss of photoreceptors is also occasionally observed, a phenotype which becomes more common with stronger alleles (not shown). The misrotation of ommatidia mutant for *RhoA* is evident early in the 3rd instar imaginal disc (during the time when tissue polarity genes are required), showing that it results from an early failure in polarity establishment (Fig. 2b). Clones of *RhoA* hypomorphic alleles in the wing also show a characteristic tissue polarity phenotype: this is manifest both in abnormal wing hair polarity, and in a multiple wing hair phenotype (Fig. 2c). Such multiple wing hair phenotypes are typical of tissue polarity genes such as *mwh* and *in*, whereas *fz* and *dsh* mutations show largely a wing hair polarity defect with only a low incidence of multiple wing hairs^{2,3}. As the multiple wing hair phenotype partially masks any underlying polarity defect, we generated a weaker loss-of-function *RhoA* state by misexpressing a dominant-negative form of the RhoA protein in part of the wing. This gives rise to a phenotype in which

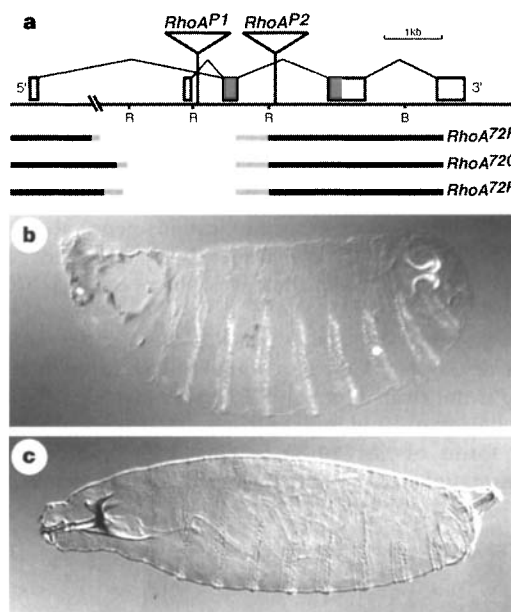


Figure 1 Organization of the *RhoA* locus, and its zygotic mutant phenotype. **a**, Genomic map of the *RhoA* locus, at polytene location 52E3-6. Exon sequences are indicated by boxes, shading shows protein coding region. Two different classes of cDNA clones were obtained with alternative first exons. Exonic structure for the two exons flanking the site of insertion of P-element *RhoA*^{P1} was determined by sequencing genomic DNA, the position of the other exons is based on Southern analysis and is only approximate. Sites of insertion of two P-elements (*RhoA*^{P1} and *RhoA*^{P2}) causing mutagenic lesions in *RhoA* are indicated above genomic DNA, restriction endonuclease sites for *Bam*HI (B) and *Eco*RI (R) are shown below. *RhoA*^{P1} and *RhoA*^{P2} were identified in a genetic screen for genes involved in early eye development from a collection of lethal P-element insertions on the 2nd chromosome²⁸, and correspond to original insertions I(2)k07236 and I(2)k07903, respectively. Further alleles of *RhoA* were generated by excision of the *RhoA*^{P1} insertion; regions of genomic DNA removed in three putative null alleles which lack the start of translation as deduced from Southern analysis are indicated (*RhoA*^{72F}, *RhoA*^{72O} and *RhoA*^{72R}). Several hypomorphic alleles were generated by excisions that did not remove protein-coding regions. The lethal phenotype of both P alleles was revertible by P-element excision and rescued by ubiquitous expression of *RhoA* cDNA clones under the control of the *tubulin* or *armadillo* promoters (not shown). **b**, Cuticle of embryo transheterozygous for two independent *RhoA* excision mutants, *RhoA*^{72F}/*RhoA*^{72O} (classified as null alleles, see above). Epidermis has failed to close in the dorsal/anterior region (with some head structures missing), giving rise to an 'anterior open' phenotype. Rarely, the dorsal epidermis also fails to close (not shown). The same defects are observed in all transheterozygous combinations of the three putative null alleles. **c**, Cuticle of wild-type embryo.

fewer cells show a multiple wing hair phenotype and a polarity defect more typical of *fz* and *dsh* is evident (Fig. 2d). We conclude that *RhoA* is required for both wing hair polarity and determination of hair number in the wing. The tissue polarity-like phenotypes of *RhoA* mutations in the eye and wing suggests that *RhoA* is a common factor required for the generation of polarity in epidermal structures.

In *RhoA* clones, which we examined in the eyes and wings, we observed only an autonomous effect, with surrounding tissue appearing phenotypically normal. This differs from the *fz* phenotype which also exhibits directional non-autonomy in both the eye and wing^{5,16}. *dsh* has been shown to function autonomously in the notum and wing^{4,17}. Similarly, clones of *dsh* in the eye show only an autonomous effect (Fig. 2e). Thus the *dsh* phenotype resembles *RhoA* more closely than *fz* in this regard.

To investigate the relationship between the function of *RhoA* and tissue polarity genes in the determination of ommatidial polarity we analysed genetic interactions. Overexpression of *Fz* in the wing causes a tissue polarity phenotype similar to *fz* loss-of-function¹⁸. As *fz* encodes a putative receptor molecule, it seems likely that this phenotype is the result of overexpression-induced ligand-independent

activation. Thus we transiently overexpressed *Fz* in a subset of photoreceptor precursors (subsequently referred to as *sev-Fz*, see Methods), and indeed found that it gave ommatidial polarity defects characterized by misrotations and incorrect chiralities (Fig. 3a). Consistent with this phenotype representing overactivation of the *Fz* signalling pathway, it was dominantly suppressed by *fz* mutations (Table 1). Genetic interactions in the wing have been interpreted to indicate that *fz* and *dsh* act in a common linear pathway⁷, and in support of this conclusion *sev-Fz* was also dominantly suppressed by *dsh* (Fig. 3b; Table 1). Remarkably, a dominant suppression of *sev-Fz* was produced by *RhoA* mutations (Fig. 3c; Table 1). In order to investigate the relationship of *RhoA* to *Dsh*, we pursued a similar strategy: *Dsh* was overexpressed in the eye (referred to as *sev-Dsh*), and this too resulted in an ommatidial polarity defect (Fig. 3e). As for *sev-Fz*, *sev-Dsh* was also dominantly suppressed by *RhoA* mutations (Fig. 3f; Table 1). Given the similarity of the mutant phenotypes of *fz*, *dsh* and *RhoA* and these genetic interactions, we propose that *RhoA*, like *Dsh*, acts downstream of *Fz* in the generation of tissue polarity.

Recent reports have implicated *Rho* family proteins in signalling to the nucleus via MAP kinase cascades, in particular via JNK/SAPK

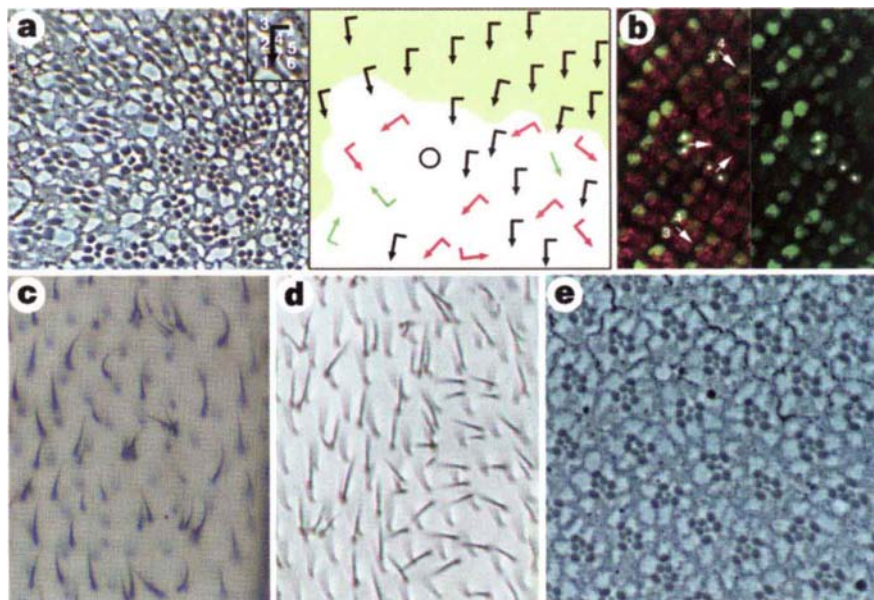


Figure 2 Phenotype of *RhoA* mutations in the eye and wing. Anterior to left in all panels; dorsal up in **a**, **b** and **e**; distal up in **c** and **d**. **a**, Section of adult eye containing homozygous mutant clone for hypomorphic excision allele *RhoA*^{72M1}, marked by absence of pigment (lower part of panel). Schematic drawing is shown on right, mutant tissue in white and adjacent wild-type tissue in yellow. Arrows indicate ommatidial orientation and chirality (see insert in left panel for orientation); black arrows: normal orientation and chirality (a $\pm 15^\circ$ wobble was allowed); red arrows: misrotated ommatidia; green arrows: ommatidia with incorrect chirality (and rotation). Circles mark unscorable ommatidia, usually due to missing photoreceptors or abnormal morphology. **b**, Immunofluorescence image of clone homozygous mutant for hypomorphic allele *RhoA*^{72BH} in a 3rd instar eye imaginal disc. Right panel: cytoplasmic green staining (anti- β -gal) marks wild-type cells, its absence indicates mutant cells, and green nuclear staining labels a subset of photoreceptors (R3/R4 strongly, and R1/R6 weakly) due to the presence of a *seven-up* enhancer trap line²⁹ (see right panel). Left panel: overlay with anti-Elav staining (red), which marks all photoreceptor nuclei, and indicates that the ommatidia in this clone have the wild-type number of photoreceptors. The R3/R4 nuclei of two wild-type ommatidia are indicated and white arrows show their normal 45° of rotation (at this stage). Note misrotation of two ommatidia containing mutant cells, evident from the pattern of *svp-lacZ* expression in R3/R4 (marked by asterisks and white arrows in left panel). **c**, Photomicrograph of an adult wing,

containing a clone homozygous mutant for hypomorphic allele *RhoA*^{72BH}. The clone is marked by the *forked* phenotype, such that all mutant cells produce a wing hair with a wavy appearance (centre of panel), whereas surrounding wild-type wing hairs are straight (left and right of panel). Wild-type cells each produce a single wing hair which points distally. Many mutant cells display a tissue polarity defect, manifest in multiple wing hairs and a failure to point distally. **d**, Photomicrograph of adult wing, showing the effect of expressing *RhoA*^{N19} in the region between veins 3 and 4 (*dpp-GAL4/UAS-RhoA*^{N19}). Note that wing hairs show polarity defects evident in their failure to point distally (in addition to frequent multiple wing hair phenotypes). **e**, Section of adult eye containing homozygous mutant *dsh*³ clone, marked by absence of pigment (lower part of panel). Ommatidia within the clone frequently show abnormal orientation and chirality, whereas those composed of wild-type tissue consistently point towards the equator (down in figure). The relatively low amounts of pigment produced by the clonal marker used do not permit identification of individual mutant cells, however, in a sample of 7 polar clonal boundaries, 28/69 ommatidia on the polar border which were not surrounded by a complete pigment lattice (and thus probably contain mutant cells) showed abnormal polarity, whereas 0/74 which were surrounded by a pigment lattice (and thus are composed of wild-type cells) showed normal polarity.

Table 1 Quantification of genetic interactions observed with *sev-Fz* and *sev-Dsh*.

(a) Suppression of <i>sev-Fz</i> /+ at 25 °C		
Genotype	Ommatidia with normal rotation or chirality (% $\pm$ s.d.)	Number of ommatidia scored
Control (+/+)	42.6 ($\pm$ 7.6)	1,125
<i>fz</i> ¹ /+	69.0 ($\pm$ 1.9)	512
<i>dsh</i> ³ /+	89.5 ($\pm$ 1.5)	335
<i>RhoA</i> ^{P1} /+	61.3 ($\pm$ 7.8)	378
<i>RhoA</i> ^{P2} /+	70.3 ($\pm$ 3.4)	360
<i>bsk</i> ¹ /+	68.2 ($\pm$ 8.1)	426
<i>bsk</i> ² /+	70.1 ($\pm$ 9.5)	349
<i>rl</i> ^{EMS698} /+	45.0 ($\pm$ 8.6)	692
<i>pk-sple</i> ¹³ /+	39.3 ($\pm$ 9.0)	588
<i>nemo</i> ^{E33} /+	37.2 ($\pm$ 3.5)	366
<i>wg</i> ^{CX4} /+	48.7 ($\pm$ 7.2)	362
(b) Suppression of <i>sev-Dsh</i> /+ at 18 °C		
Control (+/+)	32.1 ($\pm$ 3.9)	662
<i>RhoA</i> ^{P2} /+	51.0 ($\pm$ 3.1)	437
<i>RhoA</i> ^{72R} /+	59.4 ($\pm$ 5.8)	374
<i>bsk</i> ² /+	60.6 ($\pm$ 0.3)	353
<i>rl</i> ^{EMS698} /+	37.3 ($\pm$ 3.7)	372

Flies analysed were heterozygous for each of *sev-Fz* or *sev-Dsh* and the mutation of interest, from crosses grown at constant temperature. Eyes were sectioned through the equatorial region, and ommatidia in each section scored for correct or incorrect rotation/chirality relative to the equator. On average about 120 ommatidia were scored in each section, and between 3 and 8 sections from different eyes were analysed for each genotype. The genotypes shown are a selected group of many tested, most of which did not show an interaction. All modifications observed were suppressions: the figure shown is the average number of normal ommatidia per eye, with the standard deviation across all the eyes scored. Ommatidia that were rotated more than 15° away from the dorsoventral axis were scored as misrotated, whereas any ommatidium with the incorrect chiral form relative to its position dorsally or ventrally was scored as aberrant. Fewer than 1% 'unscorable' ommatidia (those missing photoreceptors or with uninterpretable arrangement of photoreceptors) were observed using the *sev-Fz* background, and 7% were observed using *sev-Dsh*.

molecules¹¹. A specific *Drosophila* MAPK homologue encoded by the *rolled* gene has been shown to act downstream of Ras in photoreceptor determination, with no apparent role in ommatidial polarity determination^{19,20}. Consistent with this, we see no genetic interactions between *rolled* (*rl*) and *sev-Fz* or *sev-Dsh*. In contrast, mutations in the gene *basket* (*bsk*), which encodes a JNK/SAPK homologue^{21,22}, suppress *sev-Fz* and *sev-Dsh* to a similar extent as *RhoA* (Fig. 3d, g; Table 1). Furthermore, the embryonic phenotypes of *RhoA* and *bsk* mutants are very similar (Fig. 1b and refs 21, 22), supporting a role for Bsk (and JNK/SAPK-like kinases) in Rho signalling. Eye clones have been generated for *bsk* hypomorphic mutations²¹, and these do not show an ommatidial polarity defect (or any other significant eye phenotype). However, the phenotype of *bsk* null clones is not known. Our results suggest an involvement for JNK/SAPK-like kinases in Fz/RhoA-mediated polarity signalling.

Two other genes required for ommatidial polarity, *pk-sple*³ and *nemo*²³, were tested for genetic interactions with *sev-Fz*. Although *pk-sple* has a similar phenotype to *fz* and *dsh*, we see no interaction between *sev-Fz* and *pk-sple* (Table 1). The *nemo* gene, which encodes a distant MAPK homologue required for correct rotation of ommatidia²³, also did not interact with *sev-Fz* (Table 1). As *nemo* mutations cause an arrest of rotation²³ (unlike *fz*, *dsh*, *pk-sple* and *RhoA* which cause random rotation), we believe that its function is required to execute rotation after Fz-signalling is complete.

Dsh also acts downstream of the signalling molecule Wingless (Wg), which can bind to the extracellular domain of Fz-like receptors and has been shown to be a ligand of the Fz-homologue

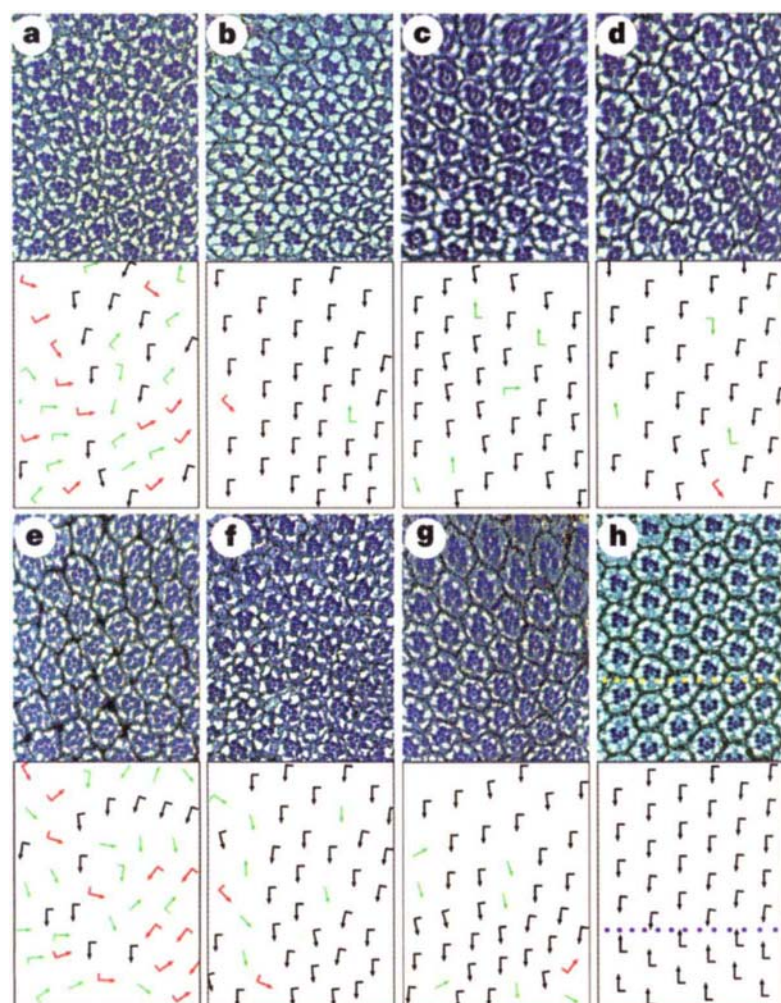


Figure 3 Genetic interactions with *sev-Fz* and *sev-Dsh* phenotypes. Anterior is left, dorsal is up in all panels. All upper panels show a section of an adult eye and the lower panels show a schematic drawing of the same area, symbols as in Fig. 2. All sections show region of eye immediately dorsal to equator, except in **h** where equator is crossed. **a-d**, Interactions with *sev-Fz*; **e-g**, Interactions with *sev-Dsh*. **a**, *sev-Fz*/+. Ommatidia are usually misrotated and many have chirality defects. **b**, *dsh*³/+; *sev-Fz*/+. Phenotype is almost completely dominantly suppressed by reduction of *dsh* dosage. **c**, *RhoA*^{P2}/+; *sev-Fz*/+. Phenotype is dominantly suppressed by reduction of *RhoA* dosage. Similar but weaker interaction was seen with *RhoA*^{P1} (see Table 1). **d**, *bsk*²/+; *sev-Fz*/+. Phenotype is dominantly suppressed by reduction of *bsk* dosage. A similar interaction was seen with *bsk*¹ (see Table 1). **e**, *sev-Dsh*/+. Ommatidia are usually misrotated and many have chirality defects. **f**, *RhoA*^{P2}/+; *sev-Dsh*/+. Phenotype is dominantly suppressed by reduction of *RhoA* dosage. Suppression is also observed with *RhoA*^{72R} (a null allele, see Table 1). **g**, *bsk*²/+; *sev-Dsh*/+. Phenotype is dominantly suppressed by reduction of *bsk* dosage. For comparison: **h** shows a wild-type eye with regular hexagonal arrangement of facets and includes the equator (marked by dotted line) towards which the ommatidia point.

Dfz²⁴. However, none of the other genes implicated in Wg signalling has a tissue polarity phenotype^{1,2} suggesting that the Wg and Fz pathways are largely distinct from each other. Consistent with this notion we did not observe an interaction between *wg* mutations and *sev-Fz* (Table 1).

In summary, our analysis of *RhoA* mutants and the genetic interactions are consistent with a role for Rho in pathways signalling to the nucleus from a membrane-bound receptor. Furthermore, there are strong parallels between the signalling pathway that we deduce for the determination of ommatidial polarity (and probably tissue polarity in general) in *Drosophila*, and the signal transduction cascades believed to function to activate SRF in mammalian cells⁹ or to activate MCM1 in response to mating pheromone in yeast⁸ (Fig. 4). In these pathways, a serpentine receptor is active at the cell surface and signals through a p21 GTPase of the Rho subfamily to a MAP kinase cascade, resulting in a transcriptional response in the nucleus. Likewise, based upon the similarities of the *fz* and *RhoA* mutant phenotypes and the genetic interactions observed, we propose that Fz signals polarity information via the *RhoA* p21 GTPase during epidermal development in *Drosophila*. □

Methods

Organization of the *RhoA* locus. DNA flanking the P-element insertion was obtained by plasmid rescue, and used to isolate genomic DNA and cDNA clones. Sequence analysis showed that they corresponded to a *RhoA* homologue previously isolated in *Drosophila* (identical in sequence, at least in the protein coding region) named *Rho1*¹⁵ and mapped to polytene location 3C. Our *in situ* hybridization experiments (chromosome localization of both P-element insertions and of flanking genomic DNA and cDNA clones) map the gene to 52E3-6 on the 2nd chromosome. As no cross-hybridization was detected at 3C, despite the identical DNA sequence of our *RhoA* clones and the published *Rho1* clones, we believe that our revised localization is correct.

Phenotypic analysis. The zygotic null embryonic phenotype was determined by crossing together adults heterozygous for two independent excision alleles which removed the start of translation, and analysing the cuticle phenotype of the progeny which failed to hatch. Homozygous mutant *RhoA* and *dsh* clones

were induced by X-ray or FLP-mediated recombination²⁵, using an *arm-lacZ* reporter²⁶ on the homologous chromosome to mark the clones. A dominant negative form of RhoA was generated by *in vitro* mutagenesis and cloned into the GAL4-inducible vector pUAST²⁷ to generate *UAS-RhoA*^{N19}. Expression of *UAS-RhoA*^{N19} in the wing was driven by *dpp-GAL4*; all attempts to drive expression in the eye resulted in early larval lethality and no adults were obtained.

Genetic interactions. The Fz and Dsh proteins were overexpressed in a subset of photoreceptors in the developing eye using *UAS-Fz* and *UAS-Dsh* lines under a *sevenless-GAL4* driver (kindly provided by P. Adler, C. Neumann and K. Basler, respectively); these combinations are referred to as *sev-Fz* and *sev-Dsh*. For genetic interactions, *sev-Fz* was raised at 25 °C and *sev-Dsh* was raised at 18 °C as in each case this gave rise to an intermediate strength phenotype. The wild-type control strain used was *w*¹¹¹⁸.

Received 26 February; accepted 2 April 1997.

- Gubb, D. & Gardia-Bellido, A. A genetic analysis of the determination of cuticular polarity during development in *Drosophila melanogaster*. *J. Embryol. Exp. Morphol.* **68**, 37–57 (1982).
- Adler, P. N. The genetic control of tissue polarity in *Drosophila*. *BioEssays* **14**, 735–741 (1992).
- Gubb, D. Genes controlling cellular polarity in *Drosophila*. *Development* **suppl.** **1993**, 269–277 (1993).
- Theisen, H. *et al.* *dishevelled* is required during wingless signalling to establish both cell polarity and cell identity. *Development* **120**, 347–360 (1994).
- Zheng, L., Zhang, J. & Carthew, R. W. *frizzled* regulates mirror-symmetric pattern formation in the *Drosophila* eye. *Development* **121**, 3045–3055 (1995).
- Vinson, C. R., Conover, S. & Adler, P. N. A *Drosophila* tissue polarity locus encodes as protein containing seven potential transmembrane domains. *Nature* **338**, 263–264 (1989).
- Krasnow, R. E., Wong, L. L. & Adler, P. N. *dishevelled* is a component of the *frizzled* signalling pathway in *Drosophila*. *Development* **121**, 4095–4102 (1995).
- Herskowitz, I. MAP kinase pathways in yeast: for mating and more. *Cell* **80**, 187–197 (1995).
- Hill, C. S., Wynne, J. & Treisman, R. The Rho family GTPases RhoA, Rac1, and CDC42Hs regulate transcriptional activation by SRF. *Cell* **81**, 1159–1170 (1995).
- Machesky, L. M. & Hall, A. Rho: a connection between membrane receptor signalling and the cytoskeleton. *Trends Cell Biol.* **6**, 304–310 (1996).
- Vojtek, A. B. & Cooper, J. A. Rho family members: activators of MAP kinase cascades. *Cell* **82**, 527–529 (1995).
- Luo, L., Liao, Y. J., Jan, L. Y. & Jan, Y. N. Distinct morphogenetic functions of similar small GTPases: *Drosophila* Drac1 is involved in axonal outgrowth and myoblast fusion. *Genes Dev.* **8**, 1787–1802 (1994).
- Eaton, S., Auvinen, P., Luo, L., Jan, Y. N. & Simons, K. CDC42 and Rac1 control different actin-dependent processes in the *Drosophila* wing disc epithelium. *J. Cell Biol.* **131**, 151–164 (1995).
- Harden, N., Loh, H. Y., Chia, W. & Lim, L. A dominant version of the small GTP-binding protein Rac disrupts cytoskeletal structures and inhibits developmental cell shape changes in *Drosophila*. *Development* **121**, 903–914 (1995).
- Hariharan, I. K. *et al.* Characterisation of Rho GTPase family homologues in *Drosophila melanogaster*: overexpressing *Rho1* in retinal cells causes a late developmental defect. *EMBO J.* **14**, 292–302 (1995).
- Vinson, C. R. & Adler, P. N. Directional non-cell autonomy and the transmission of polarity information by the *frizzled* gene of *Drosophila*. *Nature* **329**, 549–551 (1987).
- Klingensmith, J., Nusse, R. & Perrimon, N. The *Drosophila* segment polarity gene *dishevelled* encodes a novel protein required for response to the wingless signal. *Genes Dev.* **8**, 118–130 (1994).
- Krasnow, R. E. & Adler, P. N. A single *frizzled* protein has a dual role in tissue polarity *Development* **120**, 1883–1893 (1994).
- Brunner, D. *et al.* A gain of function mutation in *Drosophila* MAP kinase activates multiple receptor tyrosine kinase signaling pathways. *Cell* **76**, 875–888 (1994).
- Biggs, W. H. *et al.* The *Drosophila* *rolled* locus encodes a MAP kinase required in the sevenless signal transduction pathway. *EMBO J.* **13**, 1628–1635 (1994).
- Riesgo-Escovar, J. R., Jenni, M., Fritz, A. & Hafen, E. The *Drosophila* Jun-N-terminal kinase is required for cell morphogenesis but not for DJun-dependent cell fate specification in the eye. *Genes Dev.* **10**, 2759–2768 (1996).
- Sluss, H. K., Han, Z., Barrett, T., Davis, R. J. & Ip, Y. T. A JNK signal transduction pathway that mediates morphogenesis and an immune response in *Drosophila*. *Genes Dev.* **10**, 2745–2758 (1996).
- Choi, K.-W. & Benzer, S. Rotation of photoreceptor clusters in the developing *Drosophila* eye requires the *nemo* gene. *Cell* **78**, 125–136 (1994).
- Bhanot, P. *et al.* A new member of the *frizzled* family from *Drosophila* functions as a Wingless receptor. *Nature* **382**, 225–230 (1996).
- Xu, T. & Rubin, G. M. Analysis of genetic mosaics in developing and adult *Drosophila* tissues. *Development* **117**, 1223–1237 (1993).
- Vincent, J.-P., Girdham, C. H. & O'Farrell, P. H. A cell-autonomous, ubiquitous marker for the analysis of *Drosophila* genetic mosaics. *Dev. Biol.* **164**, 328–331 (1994).
- Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
- Török, T., Tick, G., Alvarado, M. & Kiss, I. *P-lacW* insertion mutagenesis on the second chromosome of *Drosophila melanogaster*: isolation of lethals with different overgrowth phenotypes. *Genetics* **135**, 71–80 (1993).
- Mlodzik, M., Hiromi, Y., Weber, U., Goodman, C. S. & Rubin, G. M. The *Drosophila* *seven-up* gene, a member of the steroid receptor gene superfamily, controls photoreceptor cell fates. *Cell* **60**, 211–224 (1990).
- Yanagawa, S., van Leeuwen, F., Wodarz, A., Klingensmith, J. & Nusse, R. The *Dishevelled* protein is modified by Wingless signalling *Drosophila*. *Genes Dev.* **9**, 1087–1097 (1995).

Acknowledgements. We thank I. Kiss for allowing us to screen the P-element collection, B. Dickson and E. Hafen for initial discussions and suggestions regarding the screen and A. Cyrklaff for help with chromosome hybridizations. We thank P. Adler, K. Basler, K. Cadigan, K. Choi, F. Diaz-Benjumea, S. Eaton, D. Gubb, C. Neumann, J. Roote and the Tübingen and Bloomington stock centres for fly strains and D. Bohmann, W. Brook, S. Cohen, S. Eaton and members of the Mlodzik laboratory for helpful comments on the manuscript.

Correspondence and requests for materials should be addressed to M.M. (e-mail: mlodzik@embl-heidelberg.de).

Figure 4 Model for function of Frizzled, Dsh and RhoA in the generation of tissue polarity, showing similarity to pathways proposed for activation of MCM1 in yeast pheromone signalling⁸ and activation of SRF in mammalian cells⁹. On the basis of the homology of Dsh to proteins associated with junctional complexes⁴¹⁷, and its association with membranes during Wg signalling³⁰, it seems most plausible to place it immediately downstream of the transmembrane protein Fz. As *RhoA* and *bsk* mutations dominantly suppress the overexpression-induced ligand-independent activation phenotypes of both Fz and Dsh, RhoA and JNK/SAPK are placed downstream of both these proteins, most simply in a single linear pathway. At present we have no evidence for the involvement of a G-protein in Fz signalling.

Mdm2 promotes the rapid degradation of p53

Ygal Haupt*, Ruth Maya†, Anat Kazaz† & Moshe Oren†

* Lautenberg Center for General and Tumor Immunology,

The Hebrew University Haddassah Medical School, Jerusalem, 91120, Israel

† Department of Molecular Cell Biology, The Weizmann Institute of Science, Rehovot 76100, Israel

The p53 tumour-suppressor protein exerts antiproliferative effects, including growth arrest and apoptosis, in response to various types of stress¹. The activity of p53 is abrogated by mutations that occur frequently in tumours, as well as by several viral and cellular proteins^{1,2}. The Mdm2 oncoprotein is a potent inhibitor of p53 (ref. 3). Mdm2 binds the transcriptional activation domain of p53 and blocks its ability to regulate target genes^{3,4} and to exert antiproliferative effects⁴⁻⁷. On the other hand, p53 activates the expression of the *mdm2* gene¹ in an autoregulatory feedback loop³. The interval between p53 activation and consequent Mdm2 accumulation defines a time window during which p53 exerts its effects⁸. We now report that Mdm2 also promotes the rapid degradation of p53 under conditions in which p53 is otherwise stabilized. This effect of Mdm2 requires binding of p53; moreover, a small domain of p53, encompassing the Mdm2-binding site, confers Mdm2-dependent destabilization upon heterologous proteins. Raised amounts of Mdm2 strongly repress mutant p53 accumulation in tumour-derived cells. During recovery from DNA damage, maximal Mdm2 induction coincides with rapid p53 loss. We propose that the Mdm2-promoted degradation of p53 provides a new mechanism to ensure effective termination of the p53 signal.

To investigate the biochemical consequences of p53-Mdm2 interaction, we carried out transient transfections in p53-null human lung carcinoma H1299 cells using two *mdm2* constructs. X2 encodes a full-length Mdm2 protein that binds p53, whereas ΔXM encodes a truncated Mdm2 that does not⁹. H1299 cells were transfected with an expression plasmid for wild-type (WT) human p53, either alone or with Mdm2 expression plasmids. Twenty-six hours later, the amount of p53 protein was determined by western blot analysis. Surprisingly, p53 steady-state levels were markedly reduced in cells co-transfected with *mdm2* X2 (Fig. 1a; compare lanes 2 and 3). A similar effect was observed in HeLa cells (Fig. 1a; compare lanes 6 and 7), as well as in H1299 cells transfected with mouse p53 (Y.H. and M.O., unpublished data). A comparable reduction in p53 was not caused by ΔXM (Fig. 1a, lanes 4, 8), even though the corresponding truncated Mdm2 protein is actually found in larger amounts than the full-length, X2-encoded protein in transfected H1299 cells⁶. These results indicate that Mdm2 can significantly reduce p53 levels and that binding of Mdm2 to p53 is important for this effect.

To investigate whether an interaction between p53 and Mdm2 determines the downregulation of p53, we used p53 mutants deficient in Mdm2 binding. p53Q14S19 carries a mutant Mdm2 binding domain with Gln at position 14 and Ser at 19 (ref. 10), whereas p53 $\Delta 13-52$ lacks this domain altogether; both are refractory to downregulation by Mdm2 (Fig. 1b, lanes 1-4). By contrast, tumour-derived p53 missense mutants that can still bind Mdm2 (ref. 11) are efficiently downregulated (data not shown). Thus, the effect of Mdm2 on p53 levels requires an association between these two proteins.

Two p53-Gal4 fusion proteins were used to define the minimal p53 domain sufficient for downregulation: Gal4-p53 encodes a fusion between the Gal4 DNA-binding domain and the entire wild-type p53; the protein encoded by Gal4p53(1-42) includes only the

first 42 residues of p53 (ref. 12). As shown in Fig. 1b, downregulation of Gal4p53(1-42) (lanes 7, 8) and of Gal4-p53 (lanes 5, 6) was similar to free WT p53. Hence, Mdm2 can downregulate steady-state levels of unrelated proteins, provided they are fused to p53's Mdm2-binding site. A small region of p53 containing the Mdm2-binding site is therefore necessary and sufficient for Mdm2-mediated reduction in protein levels.

Co-transfection with Mdm2 did not reduce the amount of p53 messenger RNA (Fig. 2), but caused a moderate increase (lanes 3, 4), suggesting that Mdm2 affects either the rate of synthesis or the stability of the p53 protein. We therefore did a pulse-chase analysis of p53 in transiently transfected H1299 cells. As shown in Fig. 3a, lanes 3, 5 and 7, Mdm2 induced the rapid elimination of radiolabelled p53. In contrast, there was no decrease in radiolabelled p53 in the absence of transfected Mdm2 (Fig. 3a, lanes 2, 4, 6, and Fig. 3b). Instead, p53 appeared to increase with time, which is not uncommon when using short pulses to label stable p53 (ref. 13). Transiently transfected, overexpressed p53 is stable in human cells¹⁴, so Mdm2 must affect p53 levels predominantly, if not exclusively, by

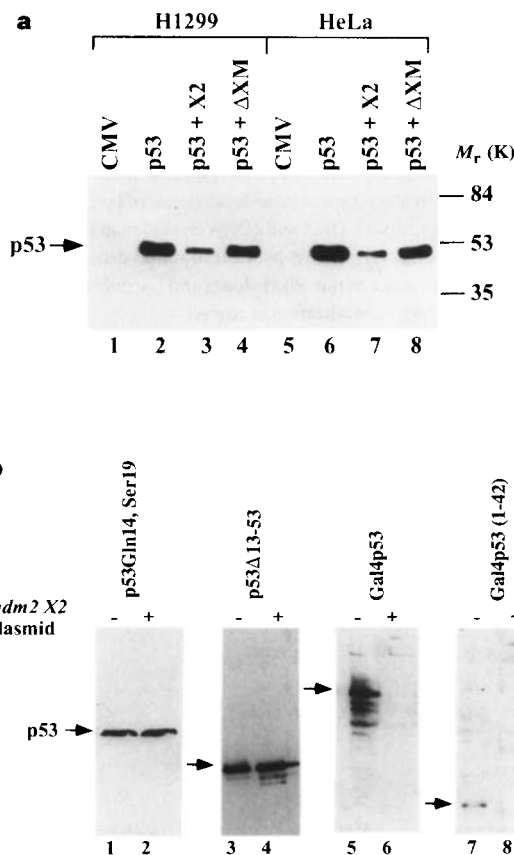


Figure 1 Mdm2 causes a reduction in steady-state levels of p53. **a**, Effects of full-length and of truncated Mdm2 on p53 levels in H1299 and HeLa cells. H1299 (lanes 1-4) or HeLa cells (lanes 5-8) were transfected transiently with the indicated plasmids. X2 encodes full-length mouse Mdm2; ΔXM encodes an Mdm2 polypeptide lacking the N-terminal p53-binding domain. CMV, vector control. 26 h later, cell extracts were prepared and western-blotted. p53 protein was detected with a mix of the monoclonal antibodies DO-1 and PAb1801. Results were similar with PAb421 antibody. Molecular size markers are indicated on the right. **b**, Effect of Mdm2 on various p53 mutants and fusion proteins. H1299 cells were transfected with the indicated p53 plasmids, either alone or with *Mdm2* X2, and p53 was analysed as in **a**. Arrows mark the expected polypeptides, identified by a mixture of DO-1 and PAb1801.

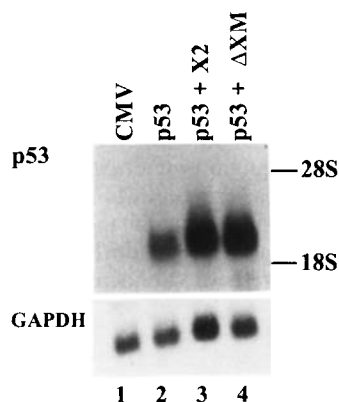
destabilizing p53 under conditions in which it is otherwise stable. This is supported by the observation that treatment of transfected H1299 cells with the proteasome inhibitor MG132 (ref. 15), which stabilizes endogenous p53 in non-transfected cells¹⁶, reverses the effect of Mdm2 (Fig. 3c). Moreover, this treatment reveals a ladder of p53 forms that migrate more slowly (Fig. 3c, lanes 2, 4). Although these bands need to be properly identified, they probably represent ubiquitinated p53, bound for proteasomal degradation. Our results indicate that Mdm2-mediated p53 degradation occurs through the ubiquitin–proteasome pathway, although a contribution from calpain^{15,17} cannot be ruled out.

These findings predict an inverse relation between Mdm2 induction and endogenous p53 levels in cells responding to stress signals. We tested this prediction in the DA-1 murine lymphoma and ML-1 myeloid leukaemia cell lines. Exposure of both lines to ionizing radiation transiently activates endogenous p53 (refs 5, 18). DA-1 cells were exposed to 200 rads of ionizing radiation, and steady-state levels of Mdm2 and p53 were determined at hourly intervals. Equal loading was monitored by probing for α -tubulin (Fig. 4c, f). As

shown in Fig. 4a, p53 protein peaked within one hour of irradiation, then decreased over the next 2 h. Mdm2 peaked 1.5–2 h postirradiation (Fig. 4b), coinciding with the marked p53 decrease. Similar results, consistent with previous observations⁵, were obtained with ML-1 cells (Fig. 4d, e) and in BALB/c-3T3 fibroblasts (data not shown). Depending on cell line and radiation dosage, a second wave of p53 accumulation was often seen several hours later (data not shown). Despite the marked changes in p53 protein levels, p53 mRNA remained fairly constant (Fig. 4g).

A ladder of slower-migrating p53 forms became apparent in ML-1 cells 3 hours after irradiation (Fig. 4d)—exactly when Mdm2 peaked and just before p53 started to decrease. The striking similarity between this and the pattern seen in transiently transfected H1299 cells (Fig. 3c) suggests that overexpressed Mdm2 targets p53 for degradation through a mechanism also used for degrading endogenous p53 in cells recovering from DNA damage.

In tumour cells bearing p53 gene mutations, the mutant protein often accumulates owing to extensive stabilization¹. It has been proposed that, unlike in normal cells, the p53 stability signal is



◀ **Figure 2** Mdm2 does not reduce p53 mRNA levels. H1299 cells were transfected with the indicated plasmids. RNA was prepared 26 h later and northern blotted with a p53 probe. Loading variations were monitored by rehybridizing the same blot with a glyceraldehyde phosphate dehydrogenase (GAPDH) probe. Ribosomal RNA size markers are shown.

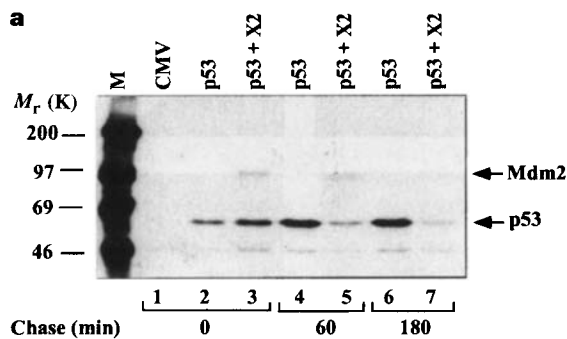
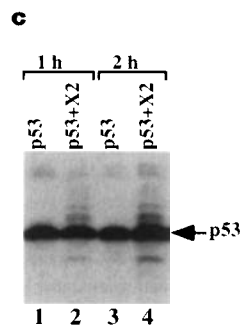
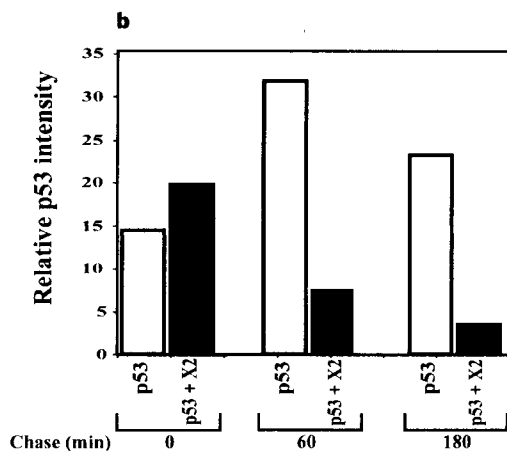


Figure 3 Mdm2 causes p53 destabilization. **a**, Pulse-chase analysis. H1299 cells transfected with various plasmids were metabolically labelled for 10 min, then chased in non-radioactive medium for the indicated periods. Samples containing equal amounts of radioactivity were immunoprecipitated with the anti-p53 monoclonal antibody PAb421 and resolved by SDS-PAGE. Molecular size markers are shown on the left. Positions of p53 and of coprecipitated Mdm2 are indicated by arrows. **b**, Quantitative analysis of pulse-chase data. The autoradiogram shown in **a** was scanned with a Phosphorimager (Fuji). Data are presented in arbitrary units. **c**, The proteasome inhibitor MG132 prevents Mdm2-mediated p53 degradation. H1299 cells were transfected as for Fig. 1a. 22 h later, MG132 was added to a final concentration of 100 μ M, and incubation resumed for an additional 1 (lanes 1, 2) or 2 (lanes 3, 4) hours. Extracts were processed as for Fig. 1a.



constitutive in such cells¹⁹. Our results indicate that this might stem from a lack of p53-mediated Mdm2 induction in cells containing only mutant p53, which is unable to activate the *mdm2* gene. Accordingly, an increased amount of Mdm2 should restore a short half-life to mutant p53. As predicted, when we transiently overexpressed Mdm2 in human T47D cells, there was a dramatic drop in endogenous mutant p53 (Fig. 5a, b), with p53 staining being reduced to near-background intensity in most cases. By contrast, a non-p53-binding form of Mdm2 was without effect (Fig. 5c, d).

Mdm2 prevents p53 from inducing growth arrest, suppressing cell transformation and inducing apoptosis^{3,6,7}, which has been attributed to inhibition by Mdm2 of p53's transcriptional activity³. Mdm2 may also oppose p53-mediated growth inhibition by acting downstream of p53 and interacting with E2F1/DP1 (ref. 21) and RB

(ref. 22) proteins. We have discovered an additional mechanism by which Mdm2 can regulate the p53 response. The extent to which p53 activity is modulated by Mdm2 may be determined by the lag period between p53 activation and p53 induction of Mdm2 expression⁸, as well as by the magnitude of this induction. Regulation of the interaction between p53 and Mdm2 depends on cell type²³. Hence the consequences of this interaction, including the ability of Mdm2 to target p53 for degradation, are also likely to vary among different cell types.

There are two possible mechanistic explanations. First, Mdm2 may contribute indirectly to rapid p53 degradation during recovery from cellular stress. Mdm2 binds within the N-terminal transactivation domain of p53 (ref. 1). DNA damage promotes increased p53 phosphorylation on specific N-terminal sites (D. Meek, personal communication). Stress-induced stabilization of p53 (ref. 24) may depend on one or more such modifications, for example at serine 15 of human p53 (ref. 25). Mdm2 can inhibit *in vitro* the phosphorylation of p53 on specific N-terminal sites (D. Meek, personal communication). Through such inhibition, Mdm2 may hinder stress-induced p53 stabilization.

Second, by analogy to HPV E6 (ref. 26), Mdm2 may directly promote p53 degradation. Covalent modification of p53's amino terminus may stabilize the protein by reducing its association with Mdm2, as shown by the interference with Mdm2 binding *in vitro* caused by phosphorylation of p53 on different N-terminal sites (C. Prives, personal communication).

One Mdm2 molecule might promote the degradation of several p53 molecules, so a molar excess of Mdm2 over p53 may not be necessary to quench the p53 response and p53 function could be blocked even when stable p53-Mdm2 complexes are barely detectable²⁷. Although we have considered the role of Mdm2 in

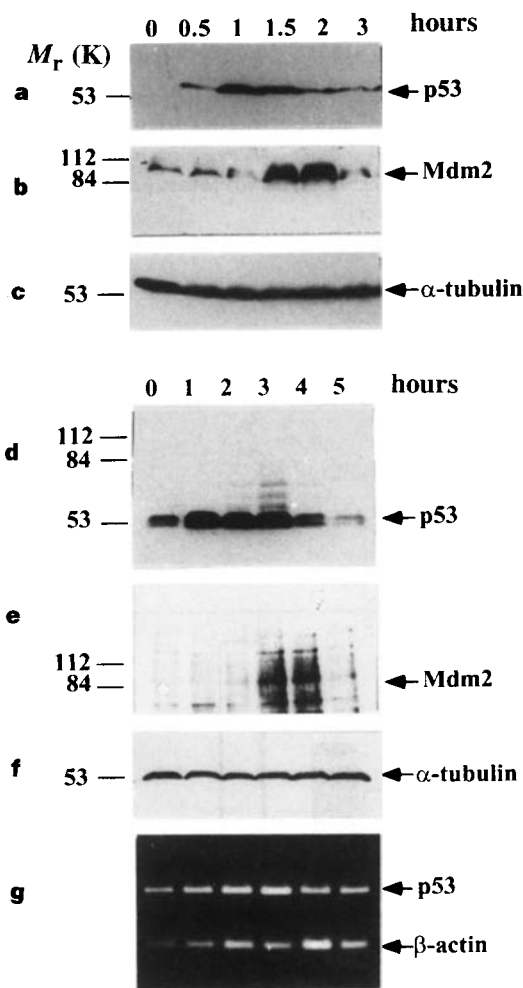


Figure 4 Accumulation of Mdm2 coincides with a rapid reduction in p53 protein levels in irradiated DA-1 and ML-1 cells. DA-1 cells were exposed to 200 Rads of ionizing radiation and then collected at the indicated times. Cell extracts were western-blotted using either a mixture of the p53-specific monoclonal antibodies PAb248 and PAb421 (**a**) or anti-mouse Mdm2 polyclonal serum⁹ (**b**). Equal loading was confirmed by reprobing with an anti- α -tubulin antibody (**c**). ML-1 cells were irradiated and analysed as described for DA-1. Levels of p53 (**d**), Mdm2 (**e**) and α -tubulin (**f**) were determined by western blot analysis using a mixture of the human p53-specific monoclonal antibodies DO-1 and PAb1801 in **d** and the anti-human Mdm2 monoclonal antibody SMP14 in **e**. RNA was prepared from other aliquots of the same cultures and semi-analysed quantitatively by RT-PCR with primers specific for human p53 (**g**). As an internal control, identical cDNA aliquots were used as PCR templates with β -actin primers.

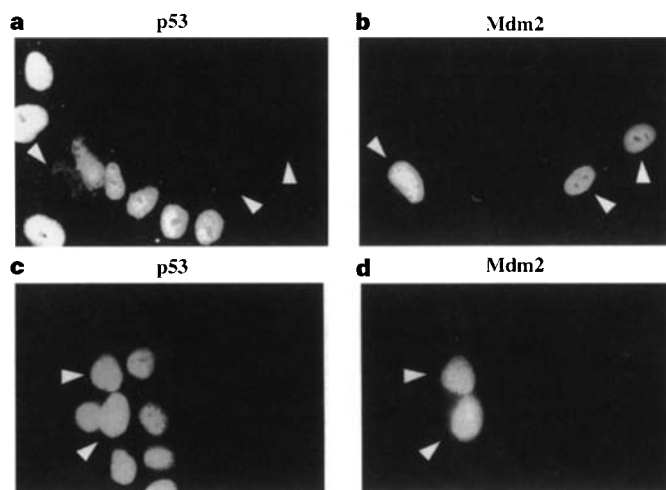


Figure 5 Mdm2 reduces mutant p53 levels in T47D cells. T47D human mammary carcinoma cells were transfected with expression plasmids X2 (**a, b**) or Δ XM (**c, d**). 26 h later, cells were fixed in methanol. Endogenous mutant p53 (**a, c**) was visualized by staining with a mixture of PAb421 and PAb1801, followed by an anti-mouse immunoglobulin Cy3-conjugated secondary antibody. Transfected mouse Mdm2 (**b, d**) was visualized with specific rabbit polyclonal serum⁹, followed by anti-rabbit FITC-conjugated secondary antibody. Each pair of adjacent panels (**a** and **b**, **c** and **d**) represents the same microscopic field, viewed under different wavelengths using a fluorescent microscope (Zeiss Axioskop). Arrowheads, cells successfully transfected and expressing exogenous Mdm2 (**b, d**). Note that in these cells expression of intact Mdm2 (**b**) reduces endogenous mutant p53 to near-background (**a**); at this relatively short exposure, residual staining can only be seen in the cell on the left in **a**. In cells in which exogenous Mdm2 is defective for p53 binding (**d**), high steady-state levels of p53 are retained (**c**). Cells were photographed at 1,000 $\times$ magnification.

regulating p53 stability during stress responses, it may also be involved in controlling basal p53 levels to an extent that depends on cell type and context. Defective Mdm2 may allow p53 to accumulate uncontrollably, explaining why p53 in the absence of Mdm2 becomes lethal to the developing embryo^{28,29}. □

Methods

Cells and transfection. 6×10^5 HeLa or H1299 cells were transfected by the calcium phosphate method using 2 μ g p53 expression plasmid and 5 μ g *mdm2* plasmid. T47D cells were transfected with lipofectamine (Gibco BRL) using 3 μ g of *mdm2* plasmid DNA. For western blotting and immunofluorescent staining, cells were collected 26 h after transfection. Expression plasmids for p53 were: pCMV-Neo-Bam-p53 (human WT p53) pRCp53Gln14Ser19 (mutant human p53)¹⁰ and p53 Δ 13–52 (mutant murine p53 lacking residues 13 to 52). Gal4-p53 fusion constructs were pGal4p53 and pGal4p53(1–42) (ref. 12) Mdm2 expression plasmids X2 and Δ XM have been described^{6,9}.

Pulse-chase analysis. H1299 cells were metabolically labelled for 10 min 26 h post-transfection as described⁶. Labelled cells were washed twice in PBS and then maintained in non-radioactive medium for 0, 60 or 180 min before collecting. Immunoprecipitation has been described⁶. p53 was immunoprecipitated with Pab421 monoclonal antibody. Precipitated proteins were resolved by SDS-PAGE and the gel fluorographed and exposed to X-ray film.

Northern blotting and RT-PCR. mRNA was prepared from transfected H1299 cells. 10 μ g mRNA was northern blotted using a p53 probe (nucleotides –25 to 656 of human p53 cDNA, radiolabelled with ³²P by random priming). For polymerase chain reaction with reverse transcription (RT-PCR), total mRNA was prepared from each sample of ML-1 cells and used as a template for reverse transcription. Amounts of cDNA were normalized by 22 cycles of PCR using specific β -actin primers (sense, 5'-GAGCACCTGTGCTGCTCACCAG-3'; antisense, 5'-GTGGTGGTGAAGCTGTAGCCACGCT-3'). Using the same amount of total cDNA, p53 cDNA was quantified by 26 cycles of PCR with primers specific for human p53 (sense, 5'-AGTGGATCCAGACTGCCTTCCGGGTCACCTG-3'; antisense, 5'-GCGGATCTAGGGCACCACCACTAT-3'). Conditions for PCR were: 94 °C for 30 s, 60 °C for 30 s, then 72 °C for 60 s.

Received 2 January; accepted 3 April 1997.

- Gottlieb, T. & Oren, M. p53 in growth control and neoplasia. *Biochim. Biophys. Acta* **1287**, 77–102 (1996).
- Shen, Y. & Shen, T. E. Viruses and apoptosis. *Curr. Opin. Gene. Dev.* **5**, 105–111 (1995).
- Picksley, S. M. & Lane, D. P. The p53–mdm2 autoregulatory feedback loop—a paradigm for the regulation of growth control by p53? *Bioessays* **15**, 689–690 (1993).
- Chen, J. D., Lin, J. Y. & Levine, A. J. Regulation of transcription functions of the p53 tumor suppressor by the *mdm-2* oncogene. *Mol. Med.* **1**, 141–142 (1995).
- Chen, C. Y. et al. Interactions between p53 and MDM2 in a mammalian cell cycle checkpoint pathway. *Proc. Natl Acad. Sci. USA* **91**, 2684–2688 (1994).
- Haupt, Y., Barak, Y. & Oren, M. Cell type-specific inhibition of p53-mediated apoptosis by *mdm2*. *EMBO J.* **15**, 1596–1606 (1996).
- Chen, J. D., Wu, X., Lin, J. Y. & Levine, A. J. *mdm-2* inhibits the G1 arrest and apoptosis functions of the p53 tumor suppressor protein. *Mol. Cell. Biol.* **16**, 2445–2452 (1996).
- Perry, M. E., Piette, J., Zawadzki, J. A., Harvey, D. & Levine, A. J. The *mdm-2* gene is induced in response to UV light in a p53-dependent manner. *Proc. Natl Acad. Sci. USA* **90**, 11623–11627 (1993).
- Barak, Y., Gottlieb, E., Juven-Gershon, T. & Oren, M. Regulation of *mdm2* expression by p53: alternative promoters produce transcripts with nonidentical translation potential. *Genes Dev.* **8**, 1739–1749 (1994).
- Lin, J. Y., Chen, J. D., Elenbaas, B. & Levine, A. J. Several hydrophobic amino acids in the p53 amino-terminal domain are required for transcriptional activation, binding to *mdm-2* and the adenovirus 5 E1B 55-kD protein. *Genes Dev.* **8**, 1235–1246 (1994).
- Hinds, P. W. et al. Mutant p53 DNA clones from human colon carcinomas cooperate with *ras* in transforming primary rat cells—a comparison of the hot spot mutant phenotype. *Cell Growth Diff.* **1**, 571–580 (1990).
- Unger, T., Nau, M. M., Segal, S. & Minna, J. D. p53—a transdominant regulator of transcription whose function is ablated by mutations occurring in human cancer. *EMBO J.* **11**, 1383–1390 (1992).
- Oren, M., Reich, N. C. & Levine, A. J. Regulation of the cellular p53 tumor antigen in teratocarcinoma cells and their differentiated progeny. *Mol. Cell. Biol.* **2**, 443–449 (1982).
- Yonish-Rouach, E. et al. Induction of apoptosis by transiently transfected metabolically stable wt p53 is transformed cell lines. *Cell Death Differ.* **1**, 39–47 (1994).
- Palombella, V. J., Rando, O. J., Goldberg, A. L. & Maniatis, T. The ubiquitin-proteasome pathway is required for processing the NF- κ B precursor protein and the activation of NF- κ B. *Cell* **78**, 773–785 (1994).
- Maki, C. J., Huibregtse, J. M. & Howley, P. M. *In vivo* ubiquitination and proteasome-mediated degradation of p53. *Cancer Res.* **56**, 2649–2654 (1996).
- Kubbutat, M. H. & Vousden, K. H. Proteolytic cleavage of human p53 by calpain: a potential regulator of protein stability. *Mol. Cell. Biol.* **17**, 460–468 (1997).
- Gottlieb, E., Lindner, S. & Oren, M. Relationship of sequence-specific transactivation and p53-regulated apoptosis in interleukin 3-dependent hematopoietic cells. *Cell Growth Diff.* **7**, 301–310 (1996).
- Lane, D. P. The regulation of p53 function: Steiner award lecture. *Int. J. Cancer* **57**, 623–627 (1994).
- Bartek, J., Iggo, R., Gannon, J. & Lane, D. P. Genetic and immunochemical analysis of mutant p53 in human breast cancer cell lines. *Oncogene* **5**, 893–899 (1990).

- Martin, K. et al. Stimulation of E2F1/DP1 transcriptional activity by MDM2 oncoprotein. *Nature* **375**, 691–694 (1995).
- Xiao, Z. X. et al. Interaction between the retinoblastoma protein and the oncoprotein Mdm2. *Nature* **375**, 694–698 (1995).
- Momand, J. & Zambetti, G. P. Analysis of the proportion of p53 bound to *mdm-2* in cells with defined growth characteristics. *Oncogene* **12**, 2279–2289 (1996).
- Kastan, M. B., Onyekwere, O., Sidransky, D., Vogelstein, B. & Craig, R. W. Participation of p53 protein in the cellular response to DNA damage. *Cancer Res.* **51**, 6304–6311 (1991).
- Fiscella, M. et al. Mutation of the serine 15 phosphorylation site of human p53 reduces the ability of p53 to inhibit cell cycle progression. *Oncogene* **8**, 1519–1528 (1993).
- Scheffner, M., Huibregtse, J. M., Vierstra, R. D. & Howley, P. M. The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. *Cell* **75**, 495–505 (1993).
- Otto, A. & Deppert, W. Upregulation of *mdm-2* expression in Meth A tumor cells tolerating wild-type p53. *Oncogene* **8**, 2591–2603 (1993).
- Jones, S. N., Roe, A. E., Donehower, L. A. & Bradley, A. Rescue of embryonic lethality in *mdm2*-deficient mice by absence of p53. *Nature* **378**, 206–208 (1995).
- Montes de Oca Luna, R., Wagner, D. S. & Lozano, G. Rescue of early embryonic lethality in *mdm2*-deficient mice by deletion of p53. *Nature* **378**, 203–206 (1995).

Acknowledgements. We thank S. Wilder and A. Peleg for technical assistance; M. B. Kastan, T. Unger, J. Chen, J. Lin and A. J. Levine for cell lines and expression plasmids; S. M. Picksley for the SMP-14 hybridoma; A. Ciechanover for MG132 and for advice; and Y. Ben Neriah for use of equipment. This work was supported in part by a grant from the National Cancer Institute, by the Minerva Foundation (Munich), and by the Leo and Julia Forchheimer Center for Molecular Genetics.

Correspondence and requests for materials should be addressed to M.O. (e-mail: lioren@dapsa1.weizmann.ac.il).

Regulation of p53 stability by Mdm2

Michael H. G. Kubbutat*, Stephen N. Jones† & Karen H. Vousden*

* ABL-Basic Research Program, NCI-FCRDC, Building 560, Room 22-96, PO Box B, Frederick, Maryland 21702-1201, USA

† Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA

The tumour-suppressor p53 is a short-lived protein that is maintained at low, often undetectable, levels in normal cells. Stabilization of the protein in response to an activating signal, such as DNA damage, results in a rapid rise in p53 levels and subsequent inhibition of cell growth¹. Tight regulation of p53 function is critical for normal cell growth and development, and one mechanism by which p53 function is controlled is through interaction with the Mdm2 protein^{2–4}. Mdm2 inhibits p53 cell-cycle arrest and apoptotic functions^{5,6} and we show here that interaction with Mdm2 can also result in a large reduction in p53 protein levels through enhanced proteasome-dependent degradation. Endogenous levels of Mdm2 are sufficient to regulate p53 stability, and overexpression of Mdm2 can reduce the amount of endogenous p53. Because *mdm2* is transcriptionally activated by p53 (refs 7, 8), this degradative pathway may contribute to the maintenance of low p53 concentrations in normal cells. Furthermore, mechanisms regulating the Mdm2-induced degradation of p53 may play a role in controlling the extent and duration of the p53 response.

Mdm2 plays a critical role in regulating p53 function, as shown by the observation that deletion of the *mdm2* gene is lethal in mice during early embryogenesis only in animals carrying a functional p53 gene^{9,10}. Simultaneous deletion of both *mdm2* and p53 gave rise to mice that developed normally, illustrating that Mdm2 is essential for negative regulation of the growth-inhibitory activities of p53 during development. Although Mdm2 is clearly not necessary for development in the absence of p53, only the amino-terminal domain of Mdm2 is necessary to inhibit p53 transcriptional and G1-arrest functions⁶, suggesting that the full-length protein may carry out additional functions. These may be related to other properties of Mdm2, such as interactions with the tumour-suppressor protein RB (ref. 11) and the transcription factor E2F (ref. 12).

We have previously described a p53 mutant, p53 Δ I, which

produces a protein lacking the amino acids constituting conserved region I (residues 13–19)¹³. Although this protein failed to interact with Mdm2 in an *in vitro* assay, it retained wild-type transcriptional activity¹³ which, in contrast to wild-type p53, is not inhibited by coexpression of Mdm2 (unpublished observations). Transient transfection assays using wild-type p53 and the p53ΔI mutant in Saos-2, a p53-null human tumour cell line, showed that p53ΔI retained the ability to induce both cell-cycle arrest and apoptosis (data not shown), indicating that the interaction with Mdm2 is not essential for these functions of p53. Previous reports have ascribed the inhibition of p53 function by Mdm2 to the physical interaction between these two proteins, which masks the transactivation domain of p53. However, during the course of these studies we noted that co-transfections of Mdm2 with wild-type p53 reproducibly reduced the number of p53-positive cells detected by flow cytometry, although this was not evident with p53ΔI (Fig. 1a). This suggested that Mdm2 may be regulating p53 expression. To examine this more closely, steady-state p53 protein levels in the transfected cells were analysed by western blotting. Wild-type p53 was expressed to lower levels than the p53ΔI mutant protein, and after cotransfection of *mdm2* the levels of wild-type p53 were markedly reduced, although this was not apparent using the p53ΔI mutant (Fig. 1b). Expression of the transfected *mdm2* in these experiments was confirmed by western blotting (data not shown). Titration of the amount of cotransfected *mdm2* plasmid showed that increasing amounts of Mdm2 correlated with decreased levels of p53 (Fig. 1c). Similar analysis of several other p53 proteins carrying mutations within the N terminus of p53 extended the correlation between the ability of p53 to bind to Mdm2 and the abrogation of expression following cotransfection of p53 and *mdm2* (Fig. 1d). The double point mutant p53 22/23, which fails to interact with Mdm2¹⁴, behaved like p53ΔI in showing constant expression levels with or without Mdm2. p53Ala15 and p53Asp15, which have substitutions of the serine at residue 15 that

is phosphorylated following activation of p53, potentially by DNA-activated protein kinase¹⁵, behaved like wild-type p53 in showing strongly reduced protein levels following cotransfection with *mdm2*. Both of these mutants also retained some ability to interact with Mdm2 *in vitro* (N. J. Marson and K.H.V., unpublished data).

To assess whether the ability of Mdm2 to modulate the amount of p53 protein expressed in cells is mediated at the transcriptional level, northern blots of RNA extracted from transiently transfected cells were carried out (Fig. 2a). This analysis showed clearly that under conditions where the p53 protein levels are dramatically reduced following cotransfection with *mdm2*, there is no discernible reduction in the p53 messenger RNA levels. Similarly, the high expression of p53ΔI protein seen in the presence of exogenous Mdm2 is not due to increased transcription of this mutant. These results indicate that regulation of p53 levels by Mdm2 is through a post-transcriptional mechanism. The levels of p53 in normal cells are regulated by the rapid turnover of the protein, and activation of the p53 pathway following DNA damage results in increased amounts of p53, in part through a lengthening of the half-life of the protein. At least one protein, human papillomavirus E6, has been shown to abrogate efficiently the normal cellular p53 response by forming an interaction with p53 and targeting it for rapid ubiquitin-dependent degradation¹⁶. To determine whether Mdm2 has a similar effect on p53 stability, we analysed the half-life of p53 protein expressed in cells with or without Mdm2 (Fig. 2b). The half-life of transfected wild-type p53 in these human cells was approximately 7.3 h, consistent with previous estimates of the half-life of endogenous p53 in human epithelial cells¹⁷, and transfection with *mdm2* resulted in the decreased stability of the wild-type p53 protein, reducing the half-life to 2.5 h. By contrast, the p53ΔI mutant protein remained stable during the 12-h chase and was not affected by coexpression of Mdm2. These results show that the interaction of Mdm2 with p53 results in an enhanced degradation of the p53 protein, thus reducing the steady-state levels of p53 detected

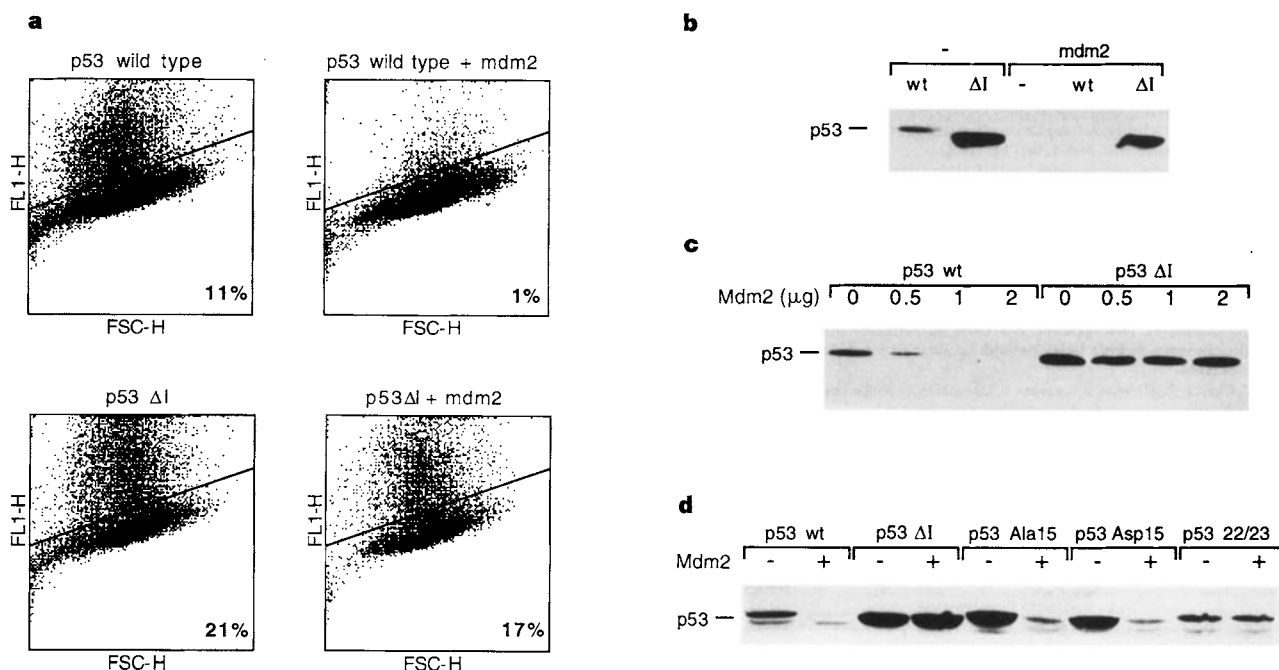


Figure 1 Mdm2 reduces expression of p53 protein. Steady-state levels of p53 protein in Saos-2 cells transiently transfected with wild-type p53 or each p53 mutant, with or without mouse *mdm2*. **a**, FACS analysis showing expression of p53 in cells following transient transfection with the indicated plasmids. Cells were labelled 24-h post-transfection with PAb421 and analysed by flow cytometry. p53-positive cells show increased fluorescence (above the line) compared with the untransfected population (below the line). The percentage of cells expressing

p53 in each population is shown. **b**, Expression of wild-type p53 and p53ΔI in the presence or absence of exogenous Mdm2. **c**, Effects of increasing amounts of Mdm2 on the levels of wild-type p53 and p53ΔI protein. Wild-type (WT) p53 and p53ΔI (1 μg), respectively, were cotransfected with the indicated amounts of *mdm2*. **d**, Levels of mutant p53 protein expression in the presence or absence of exogenous Mdm2. p53 protein was detected in lysate of transfected cells by western blot analysis using monoclonal antibody PAb1801.

by western blotting (Fig. 1).

The ability of Mdm2 to target p53 for degradation would have important implications for the mechanisms by which this regulator of p53 activity functions. The experiments described so far analysed the effects of exogenously transfected *mdm2*, and in order to determine whether physiological levels of endogenous Mdm2 were also able to carry out a similar function, we used mouse embryo fibroblasts derived from *p53*-deficient mice and from *mdm2/p53*-deficient mice^{10,18}. Analysis of wild-type p53 expression in two independently derived lines of each genotype clearly showed that the presence of endogenous Mdm2 correlates with lower levels of p53 compared to cells lacking Mdm2 (Fig. 3). The p53 Δ I mutant was expressed at equally high levels, regardless of the Mdm2 status of the cell, supporting the physiological relevance of the p53/Mdm2 interaction in the regulation of p53 stability. These results are also supported by the observation that p53 Δ I has a longer half-life than the wild-type p53 protein in human cells in the absence of exogenous Mdm2 (Fig. 2), suggesting a role for endogenous Mdm2.

Having established that endogenously expressed Mdm2 can influence p53 protein levels, we sought to demonstrate an effect of Mdm2 on endogenous p53 levels and examined both cells expressing wild-type and cells expressing mutant p53 proteins. U2OS cells, which express wild-type p53 and mount an apparently normal p53 response following DNA damage, including activation of p21^{WAF1/CIP1} and G1 arrest (data not shown), were transiently transfected with *mdm2*, DNA-damaged and collected for western-blot analysis at several time points (Fig. 4a). Endogenous p53 levels were low in these cells before damage and following DNA damage the control transfected cells showed a rapid accumulation of p53 protein over the 24-h time course. By contrast, *mdm2* overexpressing cells contained lower levels of p53 than the control cells before damage; this reduction of the already low levels of p53 protein was seen reproducibly in several experiments. Furthermore, *mdm2* expression significantly inhibited the accumulation of p53 follow-

ing DNA damage (Fig. 4a). Interestingly, analysis of Mdm2 in these cells revealed a sharp decline in the expression of the exogenous protein after 8 h, consistent with our inability to generate a stable *mdm2*-overexpressing line. Provocatively, the reduction in *mdm2* expression in these cells exactly parallels the increase in p53 protein levels. These results therefore support a role for Mdm2 in both regulating basal levels of p53 expression and modulating the DNA-damage response. We also examined the effect of Mdm2 on p53 levels in C33A cells, which express high levels of a mutant p53 protein¹⁹ that is still sensitive to degradation by Mdm2 following cotransfection in Saos-2 cells (data not shown). The p53 mutant expressed in these cells is unable to activate transcription of *mdm2*, and we considered the possibility that this inability to induce Mdm2 expression contributes to the accumulation of high levels of p53 mutant protein in these cells. We found that overexpression of wild-type Mdm2 in these cells resulted in a substantial decrease in the level of p53 protein expression in the transfected cells (Fig. 4b), an effect that was not seen following transfection of a control plasmid. A similar decrease in p53 protein expression was detected by flow cytometric analysis of C33A cells transfected with wild-type *mdm2* compared with cells transfected with an *mdm2* mutant, which encoded a protein that failed to target p53 for degradation (data not shown).

We next sought to identify the mechanism by which proteolytic degradation of p53 takes place. Several previous studies have suggested that p53 is normally ubiquitinated and degraded through the proteasome²⁰, although other proteolytic pathways may also be involved²¹. We therefore examined the effect of a specific inhibitor of the proteasome, lactacystin²², on degradation of p53 by Mdm2 (Fig. 5). Inhibition of proteasome function in *mdm2*-expressing mouse embryo fibroblasts resulted in a clear defect in the ability to degrade exogenous p53 (Fig. 4a), although this effect was not seen in cells lacking Mdm2. Treatment of human cells expressing endogenous wild-type p53 and Mdm2 with this proteasome inhibitor resulted in enhanced stability of the endogenous p53 (Fig. 5b) and an inhibition of Mdm2-targeted p53 degradation could also be seen following cotransfection of these cells with *mdm2* and p53 (Fig. 5b). These results strongly indicate that Mdm2 targets p53 for degradation through the proteasome, and suggests that Mdm2 is involved in the normal regulation of p53 stability.

Analysis of p53 mutants indicated that the regulation of p53 stability by Mdm2 is dependent on the interaction between the two proteins. Mdm2 is a large protein (491 amino acids for the human protein), of which only the N-terminal 100 amino acids are involved in the interaction with p53 (refs 4, 23). To determine whether additional Mdm2 sequences are required for the degradation of p53, we examined the effect of a previously described Mdm2 deletion mutant that retains the ability to interact with p53 and inhibit p53-induced cell-cycle arrest⁶. Although expression of wild-type Mdm2 resulted in the reduction of p53 protein levels (Fig. 6), expression of the Mdm2 mutant (Δ 222–437) not only failed to target degradation of p53, but resulted in a significant elevation of p53 levels. These results show that binding of Mdm2 to p53 is not sufficient for

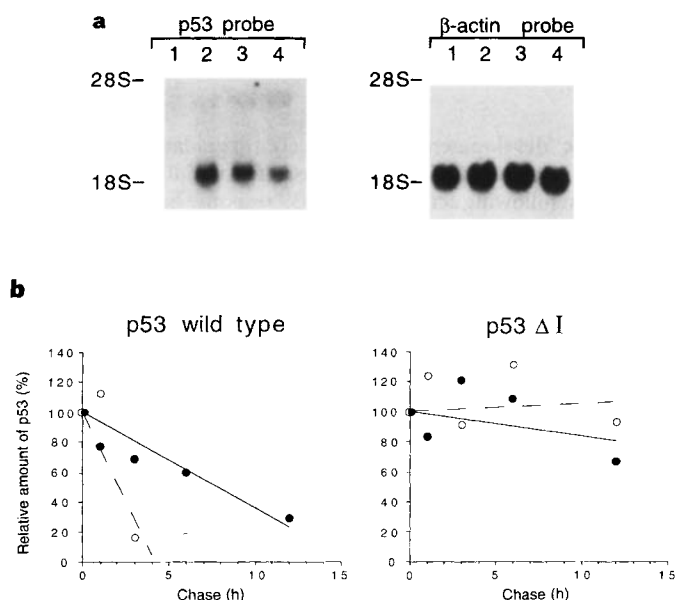


Figure 2 Mdm2 does not affect p53 transcription, but reduces p53 protein stability. **a**, Northern blot analysis of total RNA extracted from Saos-2 cells transiently transfected with vector only (1), wild-type p53 (2), wild-type p53 plus mouse *mdm2* (3) or p53 Δ I plus mouse *mdm2* (4). The filter was probed sequentially for expression of p53 and β -actin. **b**, Pulse chase of p53 proteins in transiently transfected Saos-2 cells in the presence (empty circles, dashed line) or absence (filled circles, solid line) of exogenous mouse Mdm2. Amounts of labelled p53 protein detected at each time point were calculated relative to the amount present following 0-h chase and regression was calculated using Microsoft Excel.

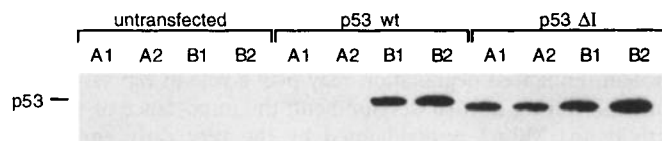


Figure 3 Expression of exogenous wild-type p53 and p53 Δ I protein in two independent clones of p53-null (A1, A2) and p53/*mdm2* double-null (B1, B2) mouse embryo fibroblasts. Equal transfection efficiency of each line was confirmed by cotransfection of a β -galactosidase expression vector (data not shown). p53 protein was detected after immunoprecipitation with PAb421 by western blot analysis using rabbit serum CM-1.

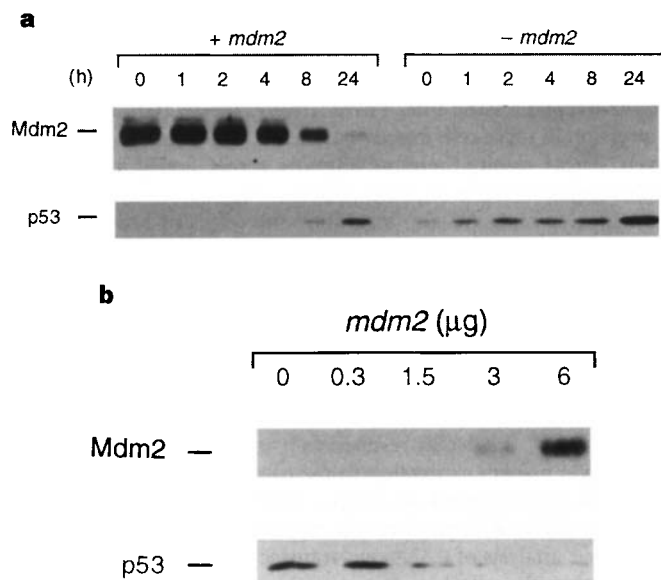


Figure 4 Mdm2 reduces endogenous p53 levels and delays DNA damage-induced p53 accumulation. **a**, U2OS cells transfected with 3 μ g human *mdm2*-expressing plasmid or empty vector were treated with 5 nM actinomycin-D to induce a p53 response. At the indicated times successfully transfected cells were isolated using the Capture-Tec kit (Clontech) and p53 and Mdm2 protein expression was detected by western blotting using monoclonal antibodies DO-1 and IF2. **b**, C33A cells positively transfected with the indicated amounts of human *mdm2*-expressing plasmid or empty vector were analysed for p53 and Mdm2 protein expression by western blotting.

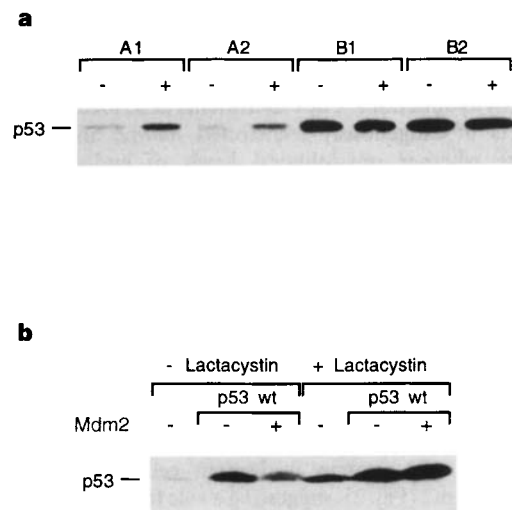


Figure 5 Inhibition of Mdm2-mediated degradation of p53 by lactacystin. **a**, Two independent clones of *p53*-null (A1, A2) and *p53/mdm2* double-null (B1, B2) mouse embryo fibroblasts were transfected with 1.5 μ g wild-type *p53* plasmid. Cells were treated 24 h after transfection with medium containing 25 μ M lactacystin (+) or left untreated. p53 protein was detected as described in Fig. 3. **b**, U2OS cells were cotransfected with wild-type *p53* plasmid and mouse *mdm2*-expressing plasmid (+) or empty vector (-). Treatment with lactacystin is described in **a**. p53 protein was detected in cell lysates by immunoblot analysis using antibody PAb1801.

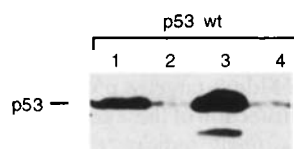


Figure 6 Binding of Mdm2 to p53 is not sufficient to induce degradation. U2OS cells were cotransfected with wild-type *p53* together with empty vector (1), human wild-type *mdm2* (2), human *mdm2* Δ 222-437 (3) encoding a mutant form of Mdm2 that can bind to p53, and mouse wild-type Mdm2 (4), respectively. p53 protein was detected in cell lysates by immunoblotting using PAb1801.

degradation and suggest a role for other domains of the Mdm2 protein. The Mdm2 Δ 222-437 mutant protein was very stably expressed, resulting in at least 20-fold more protein than the wild-type Mdm2 protein, as detected by western blotting in these cells (data not shown). Because these human cells express endogenous Mdm2, these results may indicate that at sufficiently high levels of overexpression, this mutant can act in a dominant negative manner, protecting the p53 from degradation by endogenous Mdm2 protein. Interestingly, expression of this mutant in cells expressing wild-type endogenous p53 resulted in an elevation of the normally very low p53 levels (data not shown), consistent with suggestion that the endogenous Mdm2 in these cells is contributing to p53 degradation, and that interference with wild-type Mdm2 function results in stabilization of the endogenous p53 protein.

The results reported here and in an accompanying paper²⁴ describe a new mechanism by which both mouse and human Mdm2 can regulate p53 function through degradation of the p53 protein. Enhanced degradation may play a role in repressing p53 function during normal development; the importance of such an activity for Mdm2 is highlighted by the very early embryonic lethality of Mdm2-deficient mice that retain wild-type p53. It is of interest to note that although binding of Mdm2 to p53 is apparently sufficient to inhibit transcriptional and cell-cycle arrest activities of p53, there is some evidence that this interaction may not inhibit all the apoptotic activities of the p53 protein⁵. Clearly, degradation of the p53 protein would prevent all p53-induced cell death and such a stringent inhibition of p53 function may be essential during

embryonic development. Alternatively, degradation of p53 by Mdm2 may be an efficient mechanism for rapidly reversing cell-cycle arrest following activation of a p53 response in some cell types. Although most cells with wild-type p53 show a very rapid accumulation of the protein in response to activation, in some cells the high levels of p53 are maintained²⁵, whereas in other cell types p53 levels show a transient peak and then decline, corresponding with an ability to re-enter the cell cycle²⁶. During the latter response, the rise in Mdm2 levels following p53 activation coincides with the drop in p53 levels^{24,26,27}, supporting a role for Mdm2-induced degradation of p53 during recovery from DNA damage. Normal human fibroblasts, however, show a sustained, apparently irreversible p53 response in which the increased amounts of p53 detected after DNA damage do not decline over time²⁵. Interestingly, all the characteristics of a normal p53 response are maintained in these cells, including the enhanced expression of p53-inducible genes such as *p21*^{WAF1/CIP1} and *mdm2* (S. Bates and K.H.V., unpublished data). It therefore seems likely that modification of either Mdm2 or p53, such as by phosphorylation, may protect from Mdm2-induced degradation under these conditions, thereby allowing a sustained p53 response.

The observation that expression of Mdm2, which is at least in part dependent on p53, regulates the stability of the p53 protein may provide some explanation for the enhanced stability of many p53 mutants often found in tumour cells, because these mutant proteins are transcriptionally inactive and therefore incapable of activating expression of the Mdm2 protein that would target

them for degradation. Finally, our study suggests a role for Mdm2 sequences distinct from those required for p53 binding, in targeting proteins for degradation, suggesting that other Mdm2-associated proteins, such as RB and E2F, may also be destabilized as a result of the interaction with Mdm2, or may modify the stability of p53 through their interaction with Mdm2. □

Methods

Plasmids, antibodies and proteasome inhibitor. Plasmids encoding mouse wild-type Mdm2 (pCOC Mdm2 X2)⁵, human wild-type Mdm2 (pCHDM), human mutant Mdm2 Δ 222–437 (pCHDM Δ 222–437)⁶, human wild-type p53 (pCB6 + p53Pro) and mutant p53 (pCB6 + p53 Δ 1 (ref. 13), pCB6 + Ala15, pCB6 + Asp15 (ref. 28) and pCMV p53 22/23 (ref. 14)) under the control of the CMV promoter have been described previously. p53-specific monoclonal antibodies Pab1801, D0-1 and Pab421 and the Mdm2-specific antibody IF2 were purchased from Oncogene Science. The polyclonal rabbit serum CM-1 was from Novocastra. The proteasome inhibitor lactacystin was purchased from E. J. Corey.

Cells and transfections. Saos-2 (p53 null), U2OS (p53 wild type), C33A (p53 mutant) cells, and p53-null and p53/mdm2 double-null mouse embryo fibroblasts (MEF) were maintained in DMEM supplemented with 10 or 15% fetal calf serum, respectively, and transiently transfected using the calcium phosphate precipitation method. Unless otherwise indicated, 1 μ g wild-type p53 or mutant p53 plasmid were cotransfected with 2 μ g Mdm2-encoding plasmid per 60-mm dish and cells were collected 24 h after transfection. Where appropriate, transfection efficiency was verified by cotransfection of 1.5 μ g of CMV-driven β -galactosidase expression plasmid and measurement of enzymatic β -galactosidase activity. For studies of endogenous p53 levels, cells were cotransfected with 1.5 μ g pHook and 1.5 μ g β galactosidase expression plasmid and transfected cells were separated from untransfected cells using the Capture-Tec kit (Clontech).

Protein analysis. Western blotting, immunoprecipitation²⁹ and flow cytometry³⁰ for p53 protein were carried out as previously described. Equal loading of total protein was confirmed by Ponceau S staining. For pHook-selected cells, lysates were normalized for β -galactosidase activity. The stability of p53 protein in transfected cells was determined by radioactive pulse labelling of cells for 1 h and chase in unlabelled medium for 0, 1, 3, 6 and 12 h followed by immunoprecipitation of p53 using monoclonal antibody Pab421. Quantification of labelled p53 protein was carried out using the STORM phosphorimager 860 (Molecular Dynamics).

Northern blot analysis. For northern blotting, full-length p53 cDNA was used as p53 probe; the β -actin probe was obtained from Clontech.

Received 2 January; accepted 3 April 1997.

- Bates, S. & Vousden, K. H. p53 in signalling checkpoint arrest or apoptosis. *Current Opin. Genet. Dev.* **6**, 1–7 (1996).
- Barak, Y. & Oren, M. Enhanced binding of a 95 kDa protein to p53 in cells undergoing p53-mediated growth arrest. *EMBO J.* **11**, 2115–2121 (1992).
- Momand, J., Zambetti, G. P., George, D. L. & Levine, A. J. The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. *Cell* **69**, 1237–1245 (1992).
- Chen, J. D., Marechal, V. & Levine, A. J. Mapping of the p53 and mdm-2 interaction domains. *Mol. Cell. Biol.* **13**, 4107–4114 (1993).
- Haupt, Y., Barak, Y. & Oren, M. Cell type-specific inhibition of p53-mediated apoptosis by mdm2. *EMBO J.* **15**, 1596–1606 (1996).
- Chen, J., Wu, X., Lin, J. & Levine, A. J. mdm-2 inhibits the G1 arrest and apoptosis functions of the p53 tumor suppressor protein. *Mol. Cell. Biol.* **16**, 2445–2452 (1996).
- Barak, Y., Juven, T., Haffner, R. & Oren, M. mdm-2 expression is induced by wild type p53 activity. *EMBO J.* **12**, 461–468 (1993).
- Wu, X. W., Bayle, J. H., Olson, D. & Levine, A. J. The p53 mdm-2 autoregulatory feedback loop. *Genes Dev.* **7**, 1126–1132 (1993).
- Montes de Oca Luna, R., Wagner, D. S. & Lozano, G. Rescue of early embryonic lethality in mdm2-deficient mice by absence of p53. *Nature* **378**, 203–206 (1995).
- Jones, S. N., Roe, A. E., Donehower, L. A. & Bradley, A. Rescue of embryonic lethality in Mdm2-deficient mice by absence of p53. *Nature* **378**, 206–208 (1995).
- Xiao, Z.-X. *et al.* Interaction between the retinoblastoma protein and the oncoprotein MDM2. *Nature* **375**, 694–697 (1995).
- Martin, K. *et al.* Stimulation of E2F1/DP1 transcriptional activity by MDM2 oncoprotein. *Nature* **375**, 691–694 (1995).
- Marston, N. J., Crook, T. & Vousden, K. H. Interaction of p53 with MDM2 is independent of E6 and does not mediate wild type transformation suppressor function. *Oncogene* **9**, 2707–2716 (1994).
- Lin, J., Chen, J., Elenbaas, B. & Levine, A. J. Several hydrophobic amino acids in the p53 amino-terminal domain are required for transcriptional activation, binding to mdm-2 and the adenovirus 5 E1B 55-kD protein. *Genes Dev.* **8**, 1235–1246 (1994).
- Lees-Miller, S. P., Sakaguchi, K., Ullrich, S. J., Appella, E. & Anderson, C. W. Human DNA-activated protein kinase phosphorylates serines 15 and 37 in the amino-terminal transactivation domain of human p53. *Mol. Cell. Biol.* **12**, 5041–5049 (1992).

- Scheffner, M., Werness, B. A., Huibregtse, J. M., Levine, A. J. & Howley, P. M. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell* **63**, 1129–1136 (1990).
- Hubbert, N. L., Sedman, S. A. & Schiller, J. T. Human papillomavirus type 16 E6 increases the degradation rate of p53 in human keratinocytes. *J. Virol.* **66**, 6237–6241 (1992).
- Jones, S. N. *et al.* The tumorigenic potential and cell growth characteristics of p53-deficient cells are equivalent in the presence or absence of Mdm2. *Proc. Natl Acad. Sci. USA* **93**, 14106–14111 (1996).
- Wrede, D., Tidy, J. A., Crook, T., Lane, D. & Vousden, K. H. Expression of RB and p53 proteins in HPV-positive and HPV-negative cervical carcinoma cell lines. *Mol. Carcinog.* **4**, 171–175 (1991).
- Maki, C. G., Huibregtse, J. & Howley, P. M. *In vivo* ubiquitination and proteasome-mediated degradation of p53. *Cancer Res.* **56**, 2649–2654 (1996).
- Kubbutat, M. H. G. & Vousden, K. H. Proteolytic cleavage of human p53 by calpain: a potential regulator of protein stability. *Mol. Cell. Biol.* **17**, 460–468 (1997).
- Fenteany, G. *et al.* Inhibition of proteasome activities and subunit-specific amino-terminal threonine modification by lactacystin. *Science* **268**, 726–731 (1995).
- Oliner, J. D. *et al.* Oncoprotein MDM2 conceals the activation domain of tumour suppressor p53. *Nature* **362**, 857–860 (1993).
- Haupt, Y., Maya, R., Kazan, A. & Oren, M. Mdm2 promotes the rapid degradation of p53. *Nature* **387**, 296–299 (1997).
- Di Leonardo, A., Linke, S. P., Clarkin, K. & Wahl, G. M. DNA damage triggers a prolonged p53-dependent G1 arrest and long-term induction of Cip1 in normal human fibroblasts. *Genes Dev.* **8**, 2540–2551 (1994).
- Perry, M. E., Piette, J., Zawadzki, J. A., Harvey, D. & Levine, A. J. The mdm-2 gene is induced in response to UV light in a p53-dependent manner. *Proc. Natl Acad. Sci. USA* **90**, 11623–11627 (1993).
- Chen, C. Y. *et al.* Interactions between p53 and MDM2 in a mammalian cell cycle checkpoint pathway. *Proc. Natl Acad. Sci. USA* **91**, 2684–2688 (1994).
- Marston, N. J. *Mutational Analysis of the Tumour Suppressor Protein p53*. Thesis, Univ. London (1996).
- Marston, N. J., Jenkins, J. R. & Vousden, K. H. Oligomerisation of full length p53 contributes to the interaction with mdm2 but not HPV E6. *Oncogene* **10**, 1707–1715 (1995).
- Rowan, S. *et al.* Specific loss of apoptotic but not cell cycle arrest function in a human tumour derived p53 mutant. *EMBO J.* **15**, 827–838 (1996).

Acknowledgements. We thank D. Lane, B. Vogelstein, A. Levine and M. Oren for gifts of plasmids and antibodies; M. Oren and Y. Haupt for discussion and sharing unpublished data; and S. Bates and A. Phillips for reading the manuscript. This work was supported by the National Cancer Institute, DHHS, under contract with ABL.

Correspondence and requests for materials should be addressed to K.H.V. (e-mail: vousden@ncicrf.gov).

Regulation of serotonin-2C receptor G-protein coupling by RNA editing

Colleen M. Burns, Hsin Chu, Susan M. Rueter, Linda K. Hutchinson, Hervé Canton, Elaine Sanders-Bush & Ronald B. Emeson

Department of Pharmacology, Vanderbilt University School of Medicine, Nashville, Tennessee 37232-6600, USA

The neurotransmitter serotonin (5-hydroxytryptamine, 5-HT) elicits a wide array of physiological effects by binding to several receptor subtypes. The 5-HT₂ family of receptors belongs to a large group of seven-transmembrane-spanning G-protein-coupled receptors and includes three receptor subtypes (5-HT_{2A}, 5-HT_{2B} and 5-HT_{2C}) which are linked to phospholipase C, promoting the hydrolysis of membrane phospholipids and a subsequent increase in the intracellular levels of inositol phosphates and diacylglycerol¹. Here we show that transcripts encoding the 2C subtype of serotonin receptor (5-HT_{2C}R) undergo RNA editing events in which genomically encoded adenosine residues are converted to inosines by the action of double-stranded RNA adenosine deaminase(s). Sequence analysis of complementary DNA isolates from dissected brain regions have indicated the tissue-specific expression of seven major 5-HT_{2C} receptor isoforms encoded by eleven distinct RNA species. Editing of 5-HT_{2C}R messenger RNAs alters the amino-acid coding potential of the predicted second intracellular loop of the receptor and can lead to a 10–15-fold reduction in the efficacy of the interaction between receptors and their G proteins. These observations indicate that RNA editing is a new mechanism for regulating serotonergic signal transduction and suggest that this post-transcriptional modification may be critical for modulating the different cellular functions that are mediated by other members of the G-protein-coupled receptor superfamily.

Sequence analysis of clones isolated from a rat striatum cDNA library predicted the presence of valine (GTG), serine (AGT) and valine (GTT) codons within the second intracellular loop of the 5-HT_{2C}R; instead, isoleucine (ATA), asparagine (AAT) and isoleucine (ATT) codons were found in the genomic DNA sequence at these positions (Fig. 1a). The A-to-G nucleotide discrepancies at these positions (termed sites A, B, C and D) were similar to those identified in the RNA editing of transcripts encoding the hepatitis delta virus antigenome and subunits of the α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) and kainate subtypes of ionotropic glutamate receptors²⁻⁶.

To investigate whether the nucleotide differences in 5-HT_{2C}R transcripts resulted from a similar RNA editing process, a 3.2-kilobase (kb) rat 5-HT_{2C}R genomic fragment containing exon 3 and the proximal region of the third intron was expressed transiently in the rat C6 glioma cell line (Fig. 1b, left). Primer-extension analysis of pre-mRNA derived from the transfected minigene identified A $\rightarrow$ G conversions at the A, C and D sites, showing that the A $\rightarrow$ G nucleotide discrepancies between genomic and cDNA sequences resulted from the post-transcriptional editing of 5-HT_{2C}R transcripts (Fig. 1b, right). Subsequent sequence analysis of reverse transcription-polymerase chain reaction (RT-PCR) products derived from 5-HT_{2C}R pre-mRNA transcripts confirmed the presence of similar A $\rightarrow$ G discrepancies for the B editing site (data not shown). Transient expression of 5-HT_{2C}R minigenes containing progressively smaller portions of intron 3 indicated that efficient editing could be achieved using an RNA transcript containing only

288 nucleotides of sequence information (Fig. 1b). Transcripts derived from a cDNA fragment extending from exon 3 through exon 4 (377 nucleotides) had background levels of editing at all sites, suggesting that the *cis*-active element(s) necessary for 5-HT_{2C}R editing included sequences in the proximal region of the third intron. Further analysis of pre-mRNA sequences using RNA folding algorithms⁷ identified an imperfect inverted repeat forming a putative RNA duplex between the 3' end of exon 3 and the proximal region of intron 3 (Fig. 1c). Introduction of mutations to disrupt proposed RNA base pairing abolished editing at all four sites, whereas introduction of compensatory sequence alterations restored editing to wild-type levels (data not shown). These results show that the editing of 5-HT_{2C}R transcripts requires the presence of an RNA duplex structure, similar to that previously identified for RNAs encoding AMPA and kainate glutamatergic receptor subunits^{4,8,9}.

To examine further the molecular mechanisms by which 5-HT_{2C}R transcripts are post-transcriptionally modified, an *in vitro* 5-HT_{2C}R editing system^{10,11} was developed using rat brain nuclear extracts fractionated by cation-exchange chromatography and a 5-HT_{2C}R RNA substrate (288 nucleotides; Fig. 1b) uniformly labelled with [α -³²P]adenosine 5'-triphosphate. Thin-layer chromatography of *in vitro* reaction products revealed two peaks of inosine (rather than guanosine) produced from radiolabelled 5-HT_{2C}R substrate (Fig. 2a); these activities co-eluted with two distinct peaks of double-stranded RNA (dsRNA)-adenosine deaminase activity which was detected upon incubation of fractionated extracts with

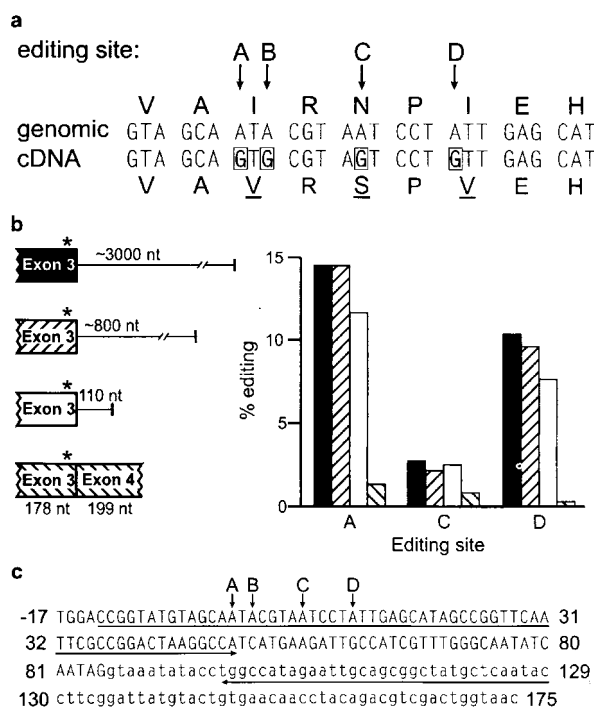


Figure 1 RNA editing of 5-HT_{2C}R transcripts. **a**, Nucleotide and predicted amino acid sequence alignments between 5-HT_{2C}R genomic and cDNA sequences; A $\rightarrow$ G nucleotide discrepancies and predicted alterations in amino acid sequence are indicated by boxed and underlined letters, respectively. The positions of four putative editing sites (A, B, C and D) are indicated. **b**, 5-HT_{2C}R minigene transcription units: editing sites (asterisk), lengths of introns (lines) and exons (boxes), and the efficiency of editing for each site upon transient transfection into rat C6 glioma cells are indicated. **c**, Nucleotide sequence of 5-HT_{2C}R genomic DNA in the proximal region of intron 3. The positions of the editing sites are indicated and an imperfect inverted repeat is shown with underlying arrows. Intronic sequence information is designated by lower-case letters and coordinates are relative to the A-site at position 0.

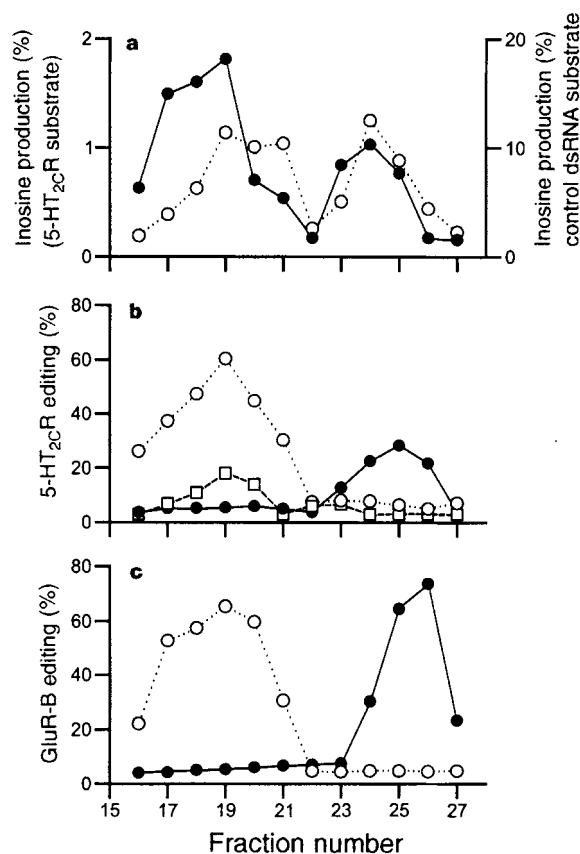


Figure 2 Chromatography of rat brain RNA-editing activity. **a**, *In vitro* production of inosine from [α -³²P]ATP labelled 5-HT_{2C}R RNA (●) or synthetic α _{2A} adrenergic receptor dsRNA (○) substrates incubated with rat brain nuclear extract fractionated by cation-exchange chromatography. **b**, Elution profile of site-specific 5-HT_{2C}R editing activity at the A (○), C (□) and D (●) sites. **c**, Elution profile of site-specific GluR-B editing activity at the +60 (○), and Q/R (●) sites.

a control dsRNA substrate. Primer-extension analysis of 5-HT_{2C}R *in vitro* reaction products identified two RNA editing activities (Fig. 2b) which co-eluted with the dsRNA-adenosine deaminase peaks (Fig. 2a). The first peak of activity was responsible for editing 5-HT_{2C}R transcripts at the A, C (Fig. 2b) and B (data not shown) sites and co-eluted with the activity mediating the modification of RNAs encoding the B-subunit of the AMPA subtype of glutamate receptor (GluR-B) at intronic position +60 (Fig. 2c). The second peak of activity specifically modified the D site of 5-HT_{2C}R transcripts (Fig. 2b) and the Q/R site of GluR-B (Fig. 2c). These results suggest that the editing of 5-HT_{2C}R transcripts *in vitro* is mediated by the actions of two separable double-stranded RNA adenosine deaminases, similar to that described for GluR-B RNAs^{11,12}, and further suggest that the editing of both 5-HT_{2C}R and GluR-B transcripts is mediated by a common cellular machinery.

Previous studies using tissue culture and *in vitro* model systems have suggested that dsRNA-specific adenosine deaminase (dsRAD or DRADA) is responsible for the modification of GluR-B transcripts at intronic position +60 (refs 12, 13), whereas dsRNA-specific editase 1 (RED1) is more selective for the Q/R site¹³. To test whether these enzymes could mediate the site-specific modification of 5-HT_{2C}R transcripts, human embryonic kidney cells (HEK293) were transiently co-transfected with a 288-nucleotide 5-HT_{2C}R minigene construct and either a rat dsRAD/DRADA or RED1 cDNA clone. Primer-extension analysis (Fig. 3a, b) of RNA transcripts derived from the transfected minigene, in the absence of exogenous dsRAD/DRADA or RED1, revealed minimal editing at all positions (Fig. 3c, d), consistent with previous observations that HEK293 cells contain a low endogenous GluR-B editing activity^{12,13}. Co-expression of dsRAD/DRADA led to a significant increase in editing efficiency at the A and C sites but had little effect upon modification of the D site. RED1 expression, by contrast, led to a significant increase in editing at all three sites, although the effects were most dramatic for the D site (Fig. 3c, d). These results show that dsRAD/DRADA and RED1 can selectively modify 5-HT_{2C}R transcripts and that these dsRNA-specific adenosine deaminases

may mediate the editing of both GluR-B and 5-HT_{2C}R pre-mRNAs. Discrepancies between the site selectivity using *in vitro* editing (Fig. 2b) rather than co-transfection (Fig. 3) strategies could result from differences in enzyme activity or from requirements for additional regulatory factors.

Although edited 5-HT_{2C}R mRNAs were initially identified in rat striatum, the potential role of dsRAD/DRADA and RED1 in 5-HT_{2C}R editing and the broad expression patterns demonstrated by these enzymes have suggested that edited 5-HT_{2C}R mRNAs could be widely distributed throughout the central nervous system^{13,14}. Primer extension and dideoxynucleotide sequence analyses of 5-HT_{2C}R mRNA derived from dissected brain regions revealed a tissue-specific pattern for these post-transcriptional modifications (Fig. 4), with the choroid plexus (the site of greatest 5-HT_{2C}R production) expressing relatively few transcripts edited at the A and

Table 1 Quantitative analysis of region-specific 5-HT_{2C}R isoform expression

Amino-acid isoforms	RNA isoforms	Whole brain	Hippocampus	Choroid plexus
VNV	ABD AD	47	36	6
VNI	AB A	18	20	0
VSV	ABCD ACD	11	22	10
VSI	ABC AC	10	18	2
INV	D	5	0	38
ISV	CD	1	0	20
INI		2	0	18

RNAs from whole rat brain, hippocampus and choroid plexus were amplified by RT-PCR and subcloned into pBSKII⁺; individual cDNA isolates were sequenced. The number of individually sequenced cDNA isolates were 100, 50 and 50 for whole brain, hippocampus and choroid plexus, respectively. 5-HT_{2C}R RNA isoforms, and the corresponding amino-acid variants predicted by them, are indicated. Numbers represent the percentage of cDNA isolates encoding each amino-acid isoform.

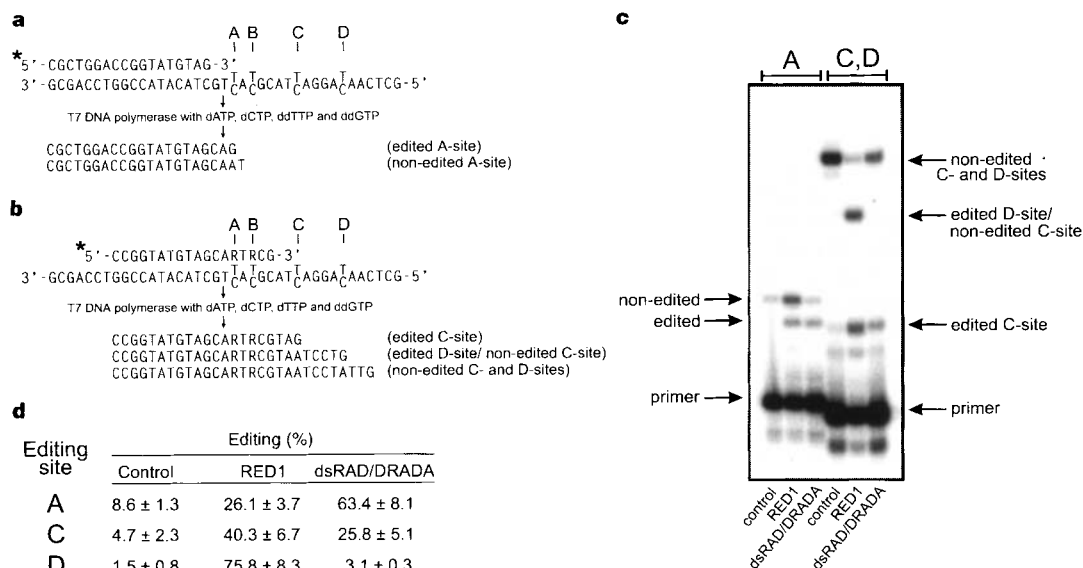


Figure 3 Co-transfection analysis of 5-HT_{2C}R transcripts and recombinant dsRNA-specific adenosine deaminases. **a**, Primer-extension strategy for quantification of editing at the 5-HT_{2C}R A-site. The nucleotide sequence of a 5'-end-labelled (asterisk) oligonucleotide sense primer, the antisense strand of the RT-PCR template, the positions of the editing sites and the sequences of expected primer-extension products are indicated. **b**, Primer-extension strategy for quantification of editing at the 5-HT_{2C}R C- and D-sites. Degenerate positions within the radiolabelled primer are indicated with an R (purine). **c**, RT-PCR/primer-extension

analysis of total RNA isolated from HEK 293 cells transiently co-transfected with a 288-bp 5-HT_{2C}R genomic construct and dsRAD/DRADA or RED1 cDNA clones. The migration positions of the primers and edited and non-edited extension products are indicated. **d**, Quantification of 5-HT_{2C}R primer-extension analysis indicating editing efficiencies at the A, C and D sites in the presence or absence of dsRAD/DRADA or RED1; values represent the mean (± standard deviation), *n* = 3.

B sites compared with other areas examined. Sequence analyses of individual cDNA isolates derived from dissected brain regions predicted the presence of seven major 5-HT_{2C}R isoforms encoded by 11 RNA species (Table 1). The most common 5-HT_{2C}R variant in whole brain and hippocampus was encoded primarily by editing at the A, B and D sites, predicting the expression of a protein isoform containing valine, asparagine and valine (VNV) at the relevant positions. The VSV, VNI and VSI variants also comprised a substantial percentage of the receptor RNAs isolated from these tissues. In contrast, the choroid plexus expressed relatively few RNA molecules encoding these isoforms owing to the reduced editing at the A and B sites (Fig. 4), and instead expressed large amounts of the INV, ISV and INI isoforms. Nine additional RNA species were also identified, but in general they represented less than 2% of the total 5-HT_{2C}R transcripts.

The observation that choroid plexus editing patterns were significantly different from those seen in other brain regions suggested that differentially edited 5-HT_{2C}R receptors may have distinct biological functions in those regions in which they are expressed. As RNA editing alters the primary amino-acid sequence in the second intracellular loop of the 5-HT_{2C}R receptor, a region implicated in receptor–G-protein (R–G) coupling¹⁵, early analyses of receptor function focused upon activation of phospholipase C and the subsequent accumulation of inositol phosphates. The fully edited VSV and unedited INI 5-HT_{2C}R isoforms were transiently transfected into NIH3T3 fibroblasts and assayed for 5-HT-stimulated inositol phosphate (InsP) accumulation. Both receptors were coupled to activation of phosphoinositide hydrolysis, but comparison of the 5-HT dose–response curve demonstrated that the agonist was 10–15-fold less potent in cells expressing the VSV isoform than the wild-type 5-HT_{2C}R (Fig. 5a). Possible explanations for this difference in 5-HT potency include alterations in the intrinsic affinity of the receptor, desensitization kinetics, spare receptors, and the efficacy of R–G coupling.

The affinity for 5-HT and (–)DOB[(–)-1-(4-bromo-2,5-dimethoxyphenyl)-2-aminopropane], determined by competition binding, was nearly equal for the INI and VSV 5-HT_{2C}R isoforms, showing that the different potency does not result from an altered agonist affinity (Fig. 5b). This idea was substantiated by antagonist competition binding, in which a series of antagonists (clozapine, ketanserin, mesulergine, bromolysergic acid diethylamide (bromo-LSD), mianserin) had similar affinities for both receptor isoforms (Fig. 5b). Desensitization of 5-HT_{2C}R receptors has also been associated with a reduced potency for 5-HT, but InsP formation in the presence of 5-HT is linear with the VSV variant for at least 30 min (data not shown), as shown for the wild-type 5-HT_{2C}R receptor¹⁶. The lower half-maximal effective concentration (EC₅₀) values determined for certain receptor isoforms (that is, the wild-type receptor) could also be explained by the presence of spare receptors, in which full receptor occupancy is not required for maximal phosphoinositide hydrolysis¹⁷. To test this, an alkylating agent (phenoxybenzamine, PBZ) was used to inactivate a fraction of the 5-HT_{2C}R receptors before analysis of 5-HT-stimulated InsP accumulation (Fig. 5c)¹⁸. The absence of spare receptors is reflected by the ability of phenoxybenzamine to decrease the maximal response of the wild-type 5-HT_{2C}R (INI) without affecting 5-HT potency. A comparable decrease in maximal response to 5-HT, with no shift in potency, was found after treatment of the VSV isoform with PBZ (data not shown). We conclude that the decreased potency of 5-HT for activating phosphoinositide hydrolysis, upon stimulation of the VSV receptor isoform, reflects reduced R–G coupling. The usual indirect methods for assessing R–G coupling, such as GTP-induced shifts in agonist affinity, agonist-promoted GTP-γS binding or GTPase activity, produce low or non-existent signals with 5-HT_{2C}R receptors and are therefore not useful for studying R–G coupling efficiency in 5-HT_{2C}R isoforms¹⁹. These methods are generally not adequate for receptors coupled to the

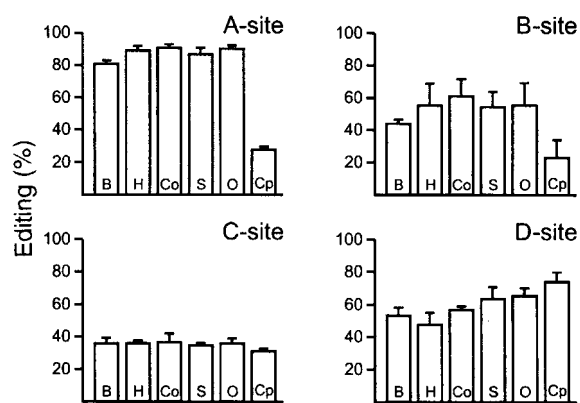


Figure 4 Analysis of site-specific 5-HT_{2C}R editing patterns. The percentage A → G conversion (editing efficiency) for each editing site (A, B, C, D) is presented for whole rat brain and dissected brain regions. Error bars represent the standard deviation from at least three independent experiments. Whole rat brain, B; hippocampus, H; cortex, Co; striatum, S; olfactory bulb, O; choroid plexus, Cp.

Gq family of G proteins and may reflect a low intrinsic GTP-turnover rate²⁰.

To define the specific amino-acid alterations responsible for decreased R–G coupling, individual 5-HT_{2C}R isoforms were transiently transfected in NIH3T3 fibroblasts and assessed for 5-HT-stimulated InsP accumulation (Fig. 5d). The results indicated that only fully edited receptor isoforms, encoding a valine, serine and valine (VSV) at positions 157, 159 and 161, respectively, exhibited reduced R–G coupling. Receptor isoforms with asparagine or isoleucine at these positions demonstrated equal 5-HT potencies as the non-edited INI receptor.

We have demonstrated that pre-mRNAs encoding the 5-HT_{2C}R receptor are post-transcriptionally modified by dsRNA-adenosine deaminase(s) to yield multiple 5-HT_{2C}R transcripts. The cellular machinery mediating these modifications is similar (or identical) to that described for RNAs encoding ionotropic glutamate receptor subunits, and may involve the enzymes dsRAD/DRADA and RED1 (refs 10–13). Region-specific editing of 5-HT_{2C}R transcripts generates multiple receptor isoforms, one of which (VSV) couples less efficiently to the intracellular signalling machinery, representing a new mechanism for regulating serotonergic signal transduction. Region- or cell-specific differences in the production of 5-HT_{2C}R isoforms may represent a stochastic event or a dynamically regulated process that modulates receptor activity thereby altering cellular responses to ambient levels of neurotransmitter. Mutagenesis of the muscarinic cholinergic receptor has shown that amino-acid residues within the conserved second intracellular loop sequence DRYXXV(I)XXPL are critical for maintaining optimal R–G interactions²¹, consistent with our evidence that editing in this region reduces the efficacy of R–G coupling. Possible explanations for this reduced R–G coupling in the fully edited 5-HT_{2C}R isoform (VSV) include altered receptor structure or receptor phosphorylation at the introduced serine moiety. Sequence analysis of mouse 5-HT_{2C}R indicates that Ser 159 may represent a potential casein kinase II site²², although amino-acid residues in this region do not strongly conform to phosphorylation consensus sequences²³. To our knowledge, the 5-HT_{2C}R is the first G-protein-coupled receptor to be regulated by RNA editing, suggesting that similar post-transcriptional modifications may modulate the different cellular responses mediated by other G-protein-coupled receptors. □

Methods

Isolation of cDNA clones. A cDNA library generated from rat striatum (λ-ZAPII; Stratagene) was screened using a randomly primed MscI–Agel fragment

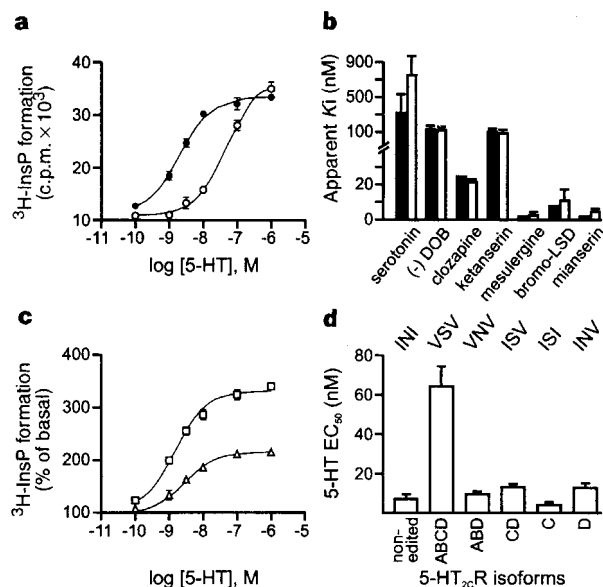


Figure 5 Functional analysis of 5-HT $_{2cR}$ isoforms. **a**, Effect of serotonin on [^3H]-InsP formation in NIH 3T3 fibroblasts transiently transfected with cDNA encoding INI (●) or VSV (○) 5-HT $_{2cR}$ isoforms; B_{max} and EC_{50} values were 5,387 versus 5,362 fmol per mg protein and 2 versus 40 nM for the INI and VSV isoforms, respectively. **b**, Apparent K_i values of agonists and antagonists for INI (filled boxes) and VSV (open boxes) 5-HT $_{2cR}$ isoforms. **c**, Effect of phenoxybenzamine (PBZ) on 5-HT-elicited phosphoinositide hydrolysis in NIH3T3 fibroblasts transiently expressing the 5-HT $_{2cR}$ (INI) isoform. Transfected cells were treated with either vehicle (□) or 1.6 μM PBZ (Δ) for 30 min before analysis. **d**, Potencies of 5-HT in eliciting [^3H]-IP accumulation in NIH3T3 cells expressing 5HT $_{2cR}$ variants. All data were analysed by nonlinear regression and values represent the mean ($\pm$ s.e.); **a**, $n = 10$; **b**, $n = 3$ –6; **c**, $n = 3$; **d**, $n = 4$ –18.

corresponding to positions 856–1,143 of the rat 5-HT $_{2cR}$ cDNA sequence as described^{12,25}. The nucleotide sequence of exon 3 was compared to the corresponding sequence of a rat 5-HT $_{2cR}$ genomic clone obtained from Genome Systems (St Louis, MO). Seven cDNA clones were isolated with various combinations of A to G discrepancies at the A, B, C and D editing sites.

A rat dsRAD/DRADA cDNA clone was obtained by screening a whole rat brain cDNA library (λ -ZAPII) with a randomly primed probe generated from a 606-bp fragment of human dsRAD/DRADA corresponding to position 1,812–2,417 as described²⁶. The isolated 4.1-kb rat cDNA clone contained the entire dsRAD/DRADA coding region with 44 and ~500 bp of 5' and 3' untranslated information, respectively.

A 2.3-kb rat RED1 cDNA clone was obtained by RT-PCR amplification of whole rat brain total RNA using sense and antisense oligonucleotide primers corresponding to positions 12–35 and 2,311–2,333 of the rat RED1 sequence¹³.

Tissue culture and transfection studies. For transfection studies (Fig. 1 (C6, rat glioma) and 3 (HEK 293, human embryonic kidney)), 5-HT $_{2cR}$ genomic fragments, as well as dsRAD/DRADA and RED1 cDNAs, were subcloned into the eukaryotic expression vector pRC/CMV (Invitrogen) and transiently expressing cells were established using calcium phosphate coprecipitation²⁵. RNAs were prepared ~60 h after transfection and editing at the A, C and D sites was analysed by primer extension. For transient transfections of 5-HT $_{2cR}$ cDNA isoforms (Fig. 5), NIH3T3 cells (90% confluence) were trypsinized, rinsed and suspended in OPTI-MEM (Gibco-BRL) at 25×10^6 cells ml^{-1} one hour before electroporation. 10–30 μg 5-HT $_{2cR}$ isoform plasmid DNA (in the eukaryotic expression vector pCMV2) was mixed with cell suspension (10×10^6 cells) in a sterile electroporation cuvette and an electric pulse (270 V, 950 mF) was applied (Bio-Rad Gene Pulser II).

Quantification of editing efficiency. Total RNA from tissue culture cells or dissected brain areas was prepared by the guanidinium-isothiocyanate method²⁵. For analysis of the transfected 5-HT $_{2cR}$ minigene constructs (Fig. 1b), reverse transcription was done with an oligonucleotide corresponding to nucleotides 180–197 of intron 3 (relative to the A editing site) or position 1,414–1,439 of the 5-HT $_{2cR}$ cDNA (exon 4)²⁴. Polymerase chain reaction (PCR) amplification was used with the initial cDNA synthesis primer and a sense primer in exon 3 (position 1,063–1,080)²⁴. Primer-extension analysis of RT-PCR reaction products was done for the A, C and D editing sites as described¹⁰. Primers for the A and C sites (Fig. 3a, b) corresponded to positions 1,136–1,153 and 1,143–1,160 (ref. 24), respectively, and were extended in the presence of either 1.2 mM deoxyadenosine 5'-triphosphate (dATP), 1.2 mM deoxycytosine 5'-triphosphate (dCTP), 5 mM dideoxyguanosine 5'-triphosphate (ddGTP) and 5 mM dideoxythymidine 5'-triphosphate (A-site) or 1.2 mM dATP, 1.2 mM dCTP, 1.2 mM deoxythymidine 5'-triphosphate and 5 mM ddGTP (C- and D-sites). For regional analysis of 5-HT $_{2cR}$ expression (Fig. 4 and Table 1), reverse transcription was done using an antisense primer in

exon 4 (positions 1,414–1,439); PCR amplification of cDNA was achieved using the initial cDNA synthesis primer and a sense primer (exon 3; positions 1,063–1,080)²⁴. PCR reaction products derived from various brain regions, representing a pool of 5-HT $_{2cR}$ cDNA isoforms, were analysed for editing by primer extension and by direct sequencing using *Taq* DNA polymerase and an antisense primer corresponding to positions 1,216–1,235 (refs 24). For the sequencing assay, products were separated on a 6% acrylamide–7M urea gel, generating a sequence ladder containing both A and G bands at positions corresponding to the A, B, C and D editing sites. The relative ratios of the A and G bands were quantified with a Molecular Dynamics phosphorimager. Four bands representing nucleotides that are not edited were chosen in both the A and G lanes and used to generate a factor for correction of differences in band intensity provided by differing deoxy/dideoxy nucleotide mixtures. Editing at the B site was consistently underestimated in this assay and sequence analyses of individual cDNA isolates (Table 1) indicated that the level of editing at the B position was similar to that seen at the A site.

In vitro editing/deamination assays. Crude rat brain nuclear extracts were prepared as described²⁷ and resuspended in buffer A (25 mM HEPES (pH 7.9), 1 mM EDTA (pH 8.0), 10% glycerol, 1 mM dithiothreitol). Extracts were fractionated by cation-exchange column chromatography (Mono S; Pharmacia) and proteins were eluted using a linear gradient from 0 to 450 mM NaCl prepared by mixing buffers A and B (buffer A plus 1 M NaCl); 1-ml fractions were collected. NaCl concentrations in individual cation-exchange fractions were determined using a conductivity meter (Radiometer; Copenhagen). The control dsRNA substrate was transcribed from a 586-bp fragment of the α_{2A} adrenergic receptor ($\alpha_{2A}\text{AR}$) subcloned into pBKSII⁺ and linearized with either *NotI* or *XbaI*. Double-stranded $\alpha_{2A}\text{AR}$ RNA was produced by annealing single-stranded transcripts as described¹⁰. For synthesis of the 5-HT $_{2cR}$ substrate, a 288-bp DNA fragment corresponding to positions –93 to +194 (relative to the A-site) was subcloned into pBKSII⁺, linearized with *XhoI*, and transcribed with T7 RNA polymerase. For GluR-B, a 646-nucleotide DNA fragment corresponding to positions –256 to +390 (relative to the Q/R site) was subcloned into pBKSII⁺, linearized with *BamHI*, and transcribed with T3 RNA polymerase¹⁰. Production of inosine (adenosine deaminase activity) was determined by TLC of *in vitro* reaction products after incubation at 30°C for 3 h in an assay (50 μl) containing 50 fmol of [α - ^{32}P]ATP labelled RNA substrate (RNA specific activity, 2.2×10^8) and 15 μl fractionated rat brain extract in buffer A as described¹⁰; the final concentration of NaCl was adjusted in each fraction to either 150 or 115 mM for the $\alpha_{2A}\text{AR}$ and 5-HT $_{2cR}$ RNA substrates, respectively. *In vitro* editing conditions were identical to those described for adenosine-deaminase assays except that the substrate RNAs were labelled with [α - ^{32}P]uridine 5'-triphosphate (RNA specific activity, 2.2×10^6) to determine RNA concentration and the reaction products were analysed by primer extension or RT-PCR sequence analyses; the final concentration of NaCl was

adjusted in each fraction to either 115 or 100 mM for the 5-HT_{2C}R and GluR-B substrates, respectively.

Pharmacological characterization. For analysis of phosphoinositide hydrolysis, transfected cells were plated in DMEM medium containing 10% calf serum, penicillin and streptomycin; 24 h after electroporation, cells were labelled overnight with 2 μ Ci [³H]myo-inositol/ml in serum-free, inositol-free DMEM. Before addition of 5-HT, cells were washed with serum-free medium and the accumulated ³H-inositol monophosphate assayed²⁸. Competition binding with [³H]mesulergine (1 mM) (Fig. 3b) was done as described²⁹ and nonspecific binding was determined with 10 μ M methysergide. Apparent K_i values were calculated according to ref. 30. Published 5-HT_{2C}R (INI) K_i values were used for clozapine, mesulergine, bromo-LSD and mianserin²⁹. 1.6 μ M PBZ decreased receptor density to 29.6% of control (Fig. 5c), as determined by saturation binding of [³H]mesulergine. EC₅₀, E_{max} and B_{max} values (n = 3) for PBZ-treated INI and VSV isoforms were 6.6 $\pm$ 2.5 nM compared with 8.5 $\pm$ 3.7 nM, 320 $\pm$ 10 compared with 240 $\pm$ 20 per cent of basal, and 3,444 $\pm$ 448 compared with 946 $\pm$ 227 fmol per mg protein, respectively.

Received 22 August 1996; accepted 7 March 1997.

- Hoyer, D. et al. IUP classification of receptors for 5-hydroxytryptamine (serotonin). *Pharmacol. Rev.* **46**, 157–203 (1994).
- Polson, A. G., Bass, B. L. & Casey, J. L. RNA editing of hepatitis delta virus antigenome by dsRNA-adenosine deaminase. *Nature* **380**, 454–456 (1996).
- Sommer, B., Kohler, M., Sprengel, R. & Seeburg, P. H. RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell* **67**, 11–19 (1991).
- Lomeli, H. et al. Control of kinetic properties of AMPA receptor channels by nuclear RNA editing. *Science* **266**, 1709–1713 (1994).
- Kohler, M., Burnashev, N., Sakmann, B. & Seeburg, P. Determinants of Ca²⁺ permeability in both TM1 and TM2 of high affinity kainate receptor channels: diversity by RNA editing. *Neuron* **10**, 491–500 (1993).
- Egebjerg, J. & Heinemann, S. F. Intron sequence directs RNA editing of the glutamate receptor subunit GluR2 coding sequence. *Proc. Natl Acad. Sci. USA* **90**, 755–759 (1993).
- Zuker, M. Computer prediction of RNA structure. *Meth. Enzymol.* **180**, 262–288 (1989).
- Higuchi, M. et al. RNA editing of AMPA receptor subunit GluR-B: a base-paired intron-exon structure determines position and efficiency. *Cell* **75**, 1361–1370 (1993).
- Herb, A., Higuchi, M., Sprengel, R. & Seeburg, P. H. Q/R site editing in kainate receptor GluR5 and GluR6 pre-mRNAs requires distant intronic sequences. *Proc. Natl Acad. Sci. USA* **93**, 1875–1880 (1996).
- Rueter, S. et al. Glutamate receptor RNA editing *in vitro* by enzymatic conversion of adenosine to inosine. *Science* **267**, 1491–1494 (1995).
- Yang, J. H., Sklar, P., Axel, R. & Maniatis, T. Editing of glutamate receptor subunit B pre-mRNA *in vitro* by site-specific deamination of adenosine. *Nature* **374**, 77–81 (1995).
- Maas, S. et al. Structural requirements for RNA editing in glutamate receptor pre-mRNAs by recombinant double-stranded RNA adenosine deaminase. *J. Biol. Chem.* **271**, 12221–12226 (1996).
- Melcher, T. et al. A mammalian RNA editing enzyme. *Nature* **379**, 460–464 (1996).
- Bass, B. RNA editing: New uses for old players in the RNA world. *The RNA World* 383–418 (Cold Spring Harbor Laboratory Press, New York, 1993).
- Gomez, J. et al. The second intracellular loop of metabotropic glutamate receptor 1 cooperates with the other intracellular domains to control coupling to G-proteins. *J. Biol. Chem.* **271**, 2199–2205 (1996).
- Westphal, R. S., Backstrom, J. R. & Sanders-Bush, E. Increased basal phosphorylation of the constitutively active serotonin 2C receptor accompanies agonist-mediated desensitization. *Mol. Pharmacol.* **48**, 200–205 (1995).
- Ariens, E. J., Beld, A. J., Rodrigues de Miranda, J. F. & Simonis, A. M. *The Receptors* 33–91 (Plenum, New York, 1979).
- Meller, E. et al. Receptor reserve for D2 dopaminergic inhibition of prolactin release *in vivo* and *in vitro*. *J. Pharmacol. Exp. Ther.* **257**, 668–675 (1991).
- Leonhardt, S., Garospe, E., Hoffman, B. J. & Teitel, M. Molecular pharmacological differences in the interaction of serotonin with 5-hydroxytryptamine 1C and 5-hydroxytryptamine 2 receptors. *Mol. Pharmacol.* **42**, 328–335 (1992).
- Ross, E. M. G protein GTPase-activating proteins: regulation of speed, amplitude, and signaling selectivity. *Rec. Prog. Horm. Res.* **50**, 207–221 (1995).
- Moro, O., Lameh, J., Högger, P. & Sadée, W. Hydrophobic amino acid in the i2 loop plays a key role in receptor–G protein coupling. *J. Biol. Chem.* **268**, 22273–22276 (1993).
- Yu, L. et al. The mouse 5-HT_{1C} receptor contains eight hydrophobic domains and is X-linked. *Mol. Brain Res.* **11**, 143–149 (1991).
- Kennelly, P. J. & Krebs, E. G. Consensus sequences as substrate specificity determinants for protein kinases and protein phosphatases. *J. Biol. Chem.* **266**, 15555–15558 (1991).
- Julius, D., MacDermott, A. B., Axel, R. & Jessell, T. M. Molecular characterization of a functional cDNA encoding the serotonin 1c receptor. *Science* **241**, 558–264 (1988).
- Ausubel, F. et al. (eds) *Current Protocols in Molecular Biology* (Wiley, New York, 1989).
- O'Connell, M. et al. Cloning of cDNAs encoding mammalian double-stranded RNA-specific adenosine deaminase. *Mol. Cell Biol.* **15**, 1389–1397 (1995).
- Gorski, K., Carneiro, M. & Schibler, U. Tissue-specific *in vitro* transcription from the mouse albumin promoter. *Cell* **47**, 767–776 (1986).
- Barker, E. L., Westphal, R. S., Schmidt, D. & Sanders-Bush, E. Constitutively active 5-hydroxytryptamine 2C receptors reveal novel inverse agonist activity of receptor ligands. *J. Biol. Chem.* **269**, 11687–11690 (1994).
- Westphal, R. S. & Sanders-Bush, E. Reciprocal binding properties of 5-hydroxytryptamine type 2C receptor agonists and inverse agonists. *Mol. Pharmacol.* **46**, 937–942 (1994).
- Cheng, Y. & Prusoff, W. H. Relationship between the inhibition constant (K_i) and the concentration of inhibitor which causes 50 per cent inhibition (IC₅₀) of enzymatic reaction. *Biochem. Pharmacol.* **22**, 3099–3108 (1973).

Acknowledgements. We thank B. Wadzinski for assistance with biochemical fractionation analysis, J. Exton, J. Barnett and J. Patton for critical reading of this manuscript, and A. Westphal and A. Poindexter for their technical expertise. This work was supported by grants from the NIH, the National Defense Medical Center, Taipei, Taiwan, and the Pharmaceutical Manufacturers Association Foundation, Inc.

Correspondence and requests for materials should be addressed to R.B.E. (e-mail: ron.emeson@mcmail.vanderbilt.edu).

Elasticity and unfolding of single molecules of the giant muscle protein titin

L. Tskhovrebova^{*†}, J. Trinick^{*}, J. A. Sleep[‡] & R. M. Simmons[‡]

^{*} Muscle Research Group, Department of Veterinary Clinical Science, Bristol University, Langford, Bristol BS18 7DY, UK

[‡] MRC Muscle & Cell Motility Unit, Randall Institute, King's College London, 26–29 Drury Lane, London WC2B 5RL, UK

The giant muscle protein titin, also called connectin, is responsible for the elasticity of relaxed striated muscle, as well as acting as the molecular scaffold for thick-filament formation^{1,2}. The titin molecule consists largely of tandem domains of the immunoglobulin and fibronectin-III types, together with specialized binding regions and a putative elastic region, the PEVK domain³. We have done mechanical experiments on single molecules of titin to determine their visco-elastic properties, using an optical-tweezers technique. On a fast (0.1s) timescale titin is elastic and force–extension data can be fitted with standard random-coil polymer models, showing that there are two main sources of elasticity: one deriving from the entropy of straightening the molecule; the other consistent with extension of the polypeptide chain in the PEVK region. On a slower timescale and above a certain force threshold, the molecule displays stress-relaxation, which occurs in rapid steps of a few piconewtons, corresponding to yielding of internal structures by about 20 nm. This stress-relaxation probably derives from unfolding of immunoglobulin and fibronectin domains.

Relaxed striated muscle possesses markedly nonlinear visco-elastic properties, including stress-relaxation and hysteresis. The steady-state force–extension relationship is at first roughly exponential, but at extreme stretch there is a yield point, after which the curve is less steep. Elasticity derives principally from connections between the ends of thick filaments and the Z-line (Fig. 1a). These connections are formed by the I-band region of the roughly 3,000K protein titin; the end-to-end length of this region increases with muscle stretch, from about zero to 1 μ m. The remaining (A-band) region of the titin molecule, 0.8 μ m long, runs to the M-line and is normally integral with the thick filament. The maximum passive (yield) force per thick filament is in the range 100–200 pN, with between 3 and 6 titin molecules in each half filament, so the maximum force per molecule lies between 15 and 70 pN. The major part of skeletal muscle titin consists of 140–160 immunoglobulin domains and 132 fibronectin-III domains; in the I-band region the 1,000–2,200 residues of the so-called PEVK region are flanked by 70–90 immunoglobulin domains, depending on isoform³. Recent studies of epitope positions in preparations of stretched muscle fibres^{4,5} suggest that titin has two physiologically important elastic mechanisms (Fig. 1b–d). At rest length, with no external force, the I-band region of the molecule probably assumes a random-coil configuration⁶ (Fig. 1b). As the sarcomere is extended force begins to rise as the molecule is straightened, with little change in secondary or tertiary structure (Fig. 1c). With further stretch, force rises more steeply as the polypeptide conformation in the PEVK region changes from a compact to a more extended configuration (Fig. 1d). At extreme lengths titin detaches from the thick filament (yield point) and the A-band part of the molecule becomes extensible⁷ (Fig. 1e). Whether the immunoglobulin or fibronectin

[†] Present address: Institute of Theoretical and Experimental Biophysics, Puschino, Moscow Region 142292, Russia.

domains can be mechanically unfolded (Fig. 1e) and contribute to extensibility or elasticity is controversial^{4,5,8-10}.

We tested these proposals by measuring the mechanical properties of individual titin molecules. We tethered a molecule between a glass surface and a polystyrene bead, using antibodies directed against epitopes located near the ends of the molecule. A bead was trapped using optical tweezers and the molecule was stretched by moving a microscope stage horizontally (Fig. 1f-j). We measured force from the displacement of the bead from the trap centre.

We designed the first type of experiment ('dynamic method') to collect data quickly, to minimize stress-relaxation. We stretched the molecule using a triangular wave movement with a frequency of about 2 Hz (Fig. 2, inset). The resulting force-extension curves were highly nonlinear (Fig. 2, continuous curve). We fitted the curves with two polymer models for a chain of contour length L : (1) the freely-jointed chain model (fjc)¹¹, in which the molecule is considered to consist of rigid segments of average length a , which are freely jointed with respect to neighbouring segments; or (2) the worm-like chain (wlc) model¹², in which the direction of the chain varies continuously, with a persistence length b (which is equivalent to half the segment length at zero external force). Assuming one component of elasticity, there were systematic deviations from the experimental curves (Fig. 2, dots), but excellent fits were obtained assuming two components in series (Fig. 2, circles). The results from both models are summarized in Table 1.

At low force most of the compliance derives from a component with an estimated contour length of 1,000–1,100 nm, with a segment length (fjc) of about 5 nm, or persistence length (wlc) also about 5 nm (component 1, Table 1, first section). The contour length corresponds to the extended length of the molecule excluding the PEVK region (Table 1, second section), and its elasticity then results from the path of the molecule adopting a random coil

Table 1 Lengths of the elastic components of titin

Method	n	L_1 (nm)	d_1 (nm)	L_2 (nm)	d_2 (nm)
Two-component models					
Freely-jointed chain	104	1,070 (125)	5.3 (0.8)	619 (122)	0.35 (0.03)
Worm-like chain	104	1,020 (58)	4.6 (0.9)	416 (45)	0.15 (0.01)
Other studies					
Sequence ^{3,4}		1,100	4	530–840	0.38
Molecular combing ²⁴		1,100		150–500	
<i>In situ</i> epitope ⁴				500–600	

The first section shows the results of fitting two-component models to the experimental force-extension curves of titin. L is the contour length and d the unitary length of a component. For the freely-jointed chain model d is the segment length, and for the worm-like chain model d is the persistence length. Figures in brackets are s.e.m. Variability in the results may be due to a proportion of the beads having multiple attachments or to tangling or knotting of the titin. The second section shows published results from studies using other methods, where L_1 is the length of the molecule, after subtracting 200 nm for the length excluded by the epitopes in the present study. L_2 is the length of the PEVK region whose size is estimated from its sequence^{3,4} and which correlates to a selectively extensible region identified by molecular combing²⁴ or epitope monitoring *in situ*⁴.

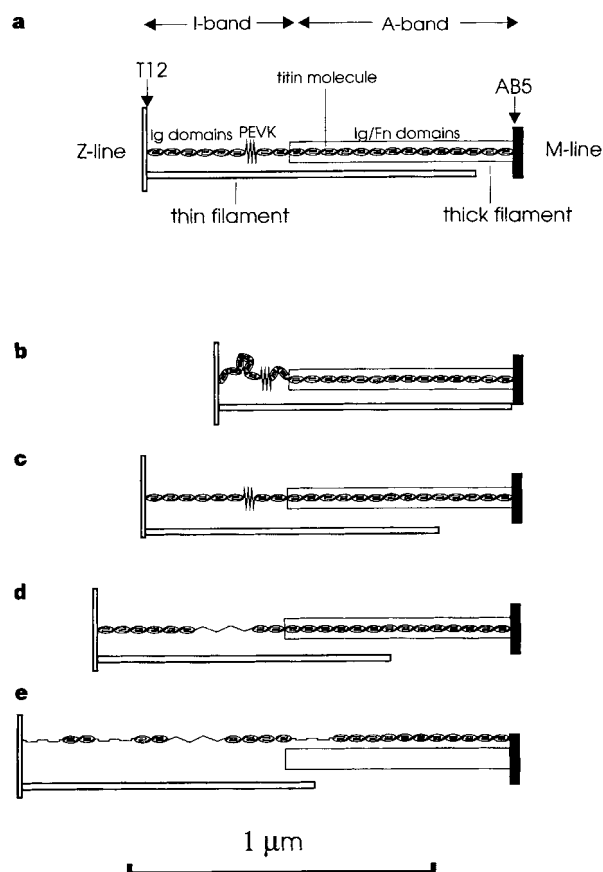
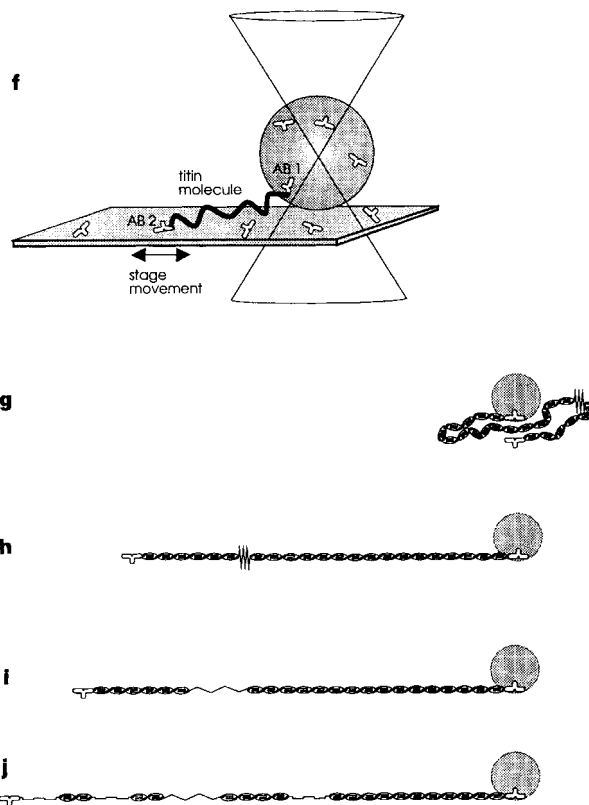


Figure 1 Schematic diagram of mechanisms of elasticity and extensibility of titin *in situ* (a-e) and in isolated molecules (f-j). **a**, Structure and location of titin. Ovals represent immunoglobulin and fibronectin domains; there is a total of about 280. Vertical arrows indicate the positions of the antibodies used. **b**, At zero stretch, the titin configuration in the I-band is a random coil and force is zero. **c**, Stretch



initially straightens the molecule and force rises. **d**, At moderate stretch, the PEVK region extends. **e**, At extreme stretch, titin detaches from the thick filament. Immunoglobulin and fibronectin domains may also unfold. **f**, Experimental procedure for single molecules. **g-j**, Mechanisms of elasticity in single-molecule experiments, corresponding to **b-e**.

configuration, with flexibility of 1–2 domains between units. The second component accounts for a greater share of the compliance at higher force (component 2, Table 1, first section). Its contour length is commensurate with the extended length of the PEVK region (Table 1, second section), and its segment length of 0.35 nm is consistent with the predicted repeat per residue of 0.38 nm of fully extended polypeptide.

In the second type of experiment ('static method') we measured the steady-state force–extension relationship by increasing the length of a molecule in steps of about 250 nm, waiting at each stage until stress-relaxation (time constant about 1 s) was nearly complete (inset to Fig. 3a). For small stretches the force–extension behaviour followed the dynamic curve. Stress-relaxation became apparent at some threshold value of force (15–60 pN), after which

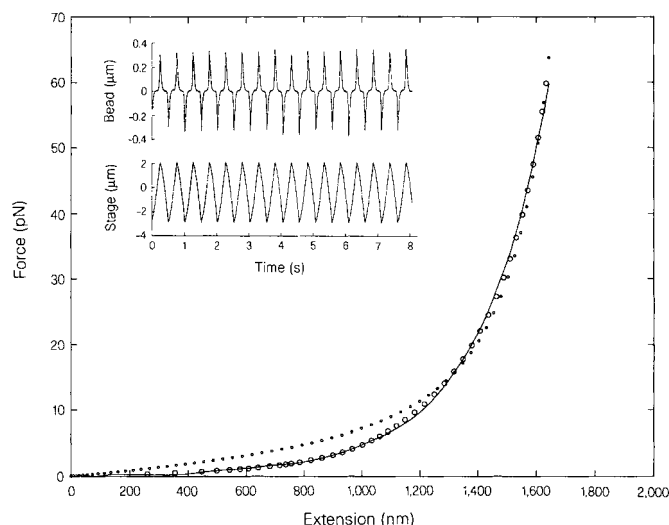


Figure 2 Dynamic force–extension relationship of titin. Experimental curve from one experiment (continuous curve); fitted using single-component worm-chain model, with $L = 1,932$ nm, $b = 0.76$ nm (dots); fitted using two-component worm-chain model, with $L_1 = 1,174$ nm, $b_1 = 4.15$ nm, $L_2 = 791$ nm, $b_2 = 0.19$ nm (circles). Fits using the freely-jointed chain model were slightly worse for a single component and equally good for two components. Inset, records of stage movement and resulting bead movement. Extension of the molecule was obtained from the stage movement after subtracting the bead movement and applying a geometric correction; force was obtained from the bead movement, trap stiffness and a geometrical correction (see Methods).

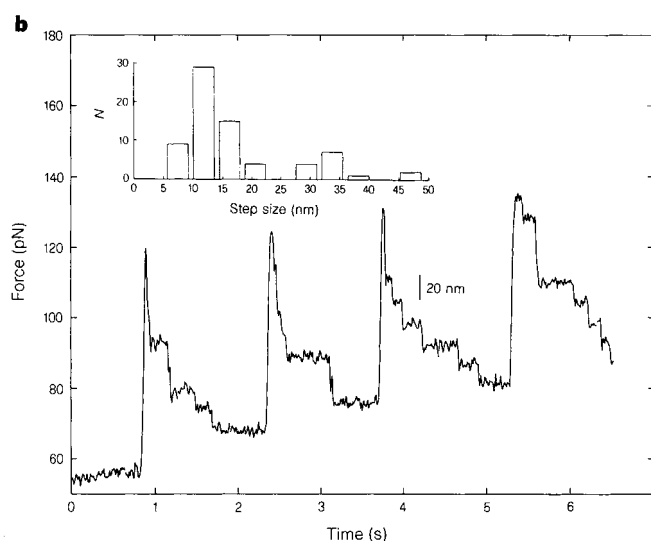
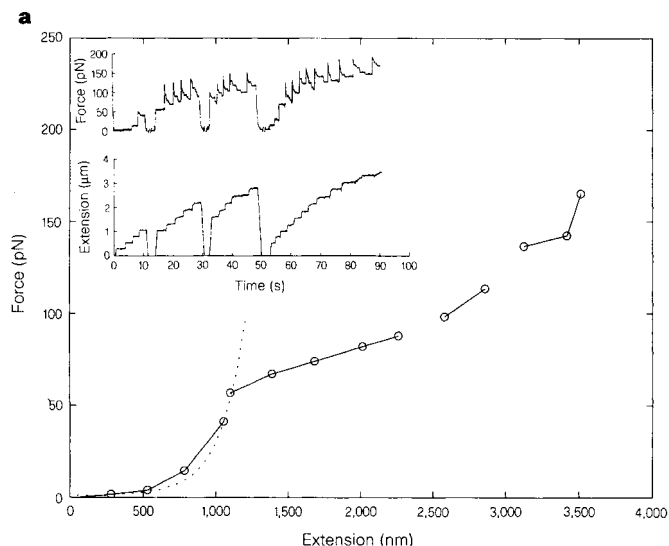
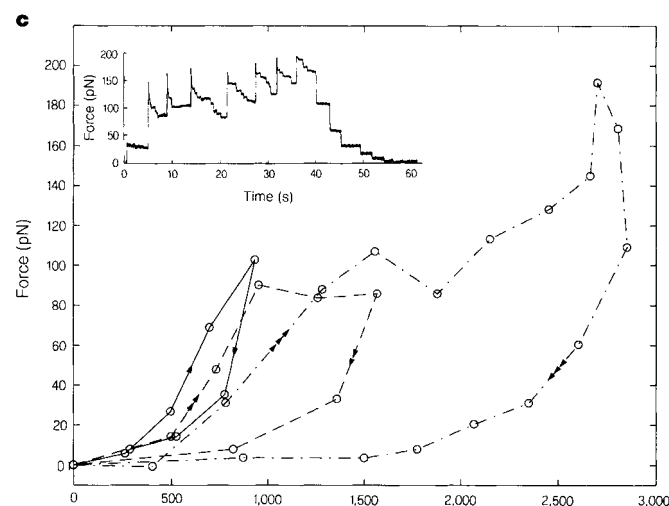


Figure 3 Static force–extension relationship. **a**, Steady-state force–extension relationship (solid lines and circles) derived from three cycles of stretches and releases (records shown in inset). For clarity, only the later non-overlapping points from succeeding cycles are shown. After the end of the last cycle the bead escaped from the trap. The dashed line shows the dynamic force–extension relationship. **b**, Stepwise stress-relaxation (enlargement of part of the record in **a**). The bar shows the scale for the corresponding extension of the molecule, obtained from the slope of the instantaneous force–extension relationship during the rapid length changes. Inset, histogram of step size; the average unitary step size in this experiment was 13 nm, but there was considerable variation between experiments, and the average for all was 19 nm. **c**, Hysteresis. Force–extension relationship from another experiment; complete curves were obtained in three stretch–release cycles. Arrows mark progression around each loop. Inset, force record from third cycle to show unchanging force levels after releases.



the curve was less steep. The average maximum extension achieved in all experiments was 3.2 μm at a force of 110 pN; 5 molecules (out of 22) were extended to 4–7 μm . Indirect evidence for such large extensions has been reported previously¹³ and must entail unfolding of immunoglobulin and/or fibronectin domains.

Stress-relaxation occurred in stepwise fashion (Fig. 3b), with a unitary drop in force in the range 3 to 10 pN (or multiples of the unitary value). The corresponding internal extension ('step size'), calculated from the change in force divided by the instantaneous stiffness, was 19 ± 11 nm (mean $\pm$ s.d., $n = 821$) from all experiments. The contour length of an unfolded domain was about 35 nm (see Methods), agreeing well with the theoretical maximum of 38 nm for domains of 100 residues.

Complete stretch-release cycles show hysteresis loops, which become increasingly wide as the degree of stretch is increased (Fig. 3c), as observed in muscle^{7,14}. There was no visible sign of reversal of stress-relaxation following length decrements (inset to Fig. 3c). However, after a completed cycle the subsequent ascending curve usually showed a partial recovery from the previous descending curve, and this was accompanied by stepwise stress-relaxation starting at about the same threshold of force as in the previous cycle (inset to Fig. 3a). This suggests that refolding occurs, but only when the external force has been reduced to a low level. Whether complete recovery can occur after a sufficiently long interval at low force remains to be investigated.

Why do domains unfold? The unperturbed equilibrium constant (K_u) for the folded-unfolded transition determined for a low-stability titin immunoglobulin domain ($\Delta G = 2.6$ kcal mol⁻¹) is about 750 (ref. 15). The half-time of unfolding of an immunoglobulin domain from the related protein twitchin ($\Delta G = 4$ kcal mol⁻¹) is 40 min and the refolding half-time is about 0.5 s (ref. 16). The energy of stretching a domain at a force of 50 pN can be calculated from our experiments to be about 1.8 kcal mol⁻¹. The strain in a domain at 50 pN is about 0.5 nm, presumably sufficient to break inter-chain bonds in the β -sheets of the domains. Allowing for the differences in stability of the titin and twitchin domains, the strain energy could reduce the time constant for unfolding to 1 s, and K_u to 4. In our experiments stretching an unfolded domain to 20 nm will increase its free energy by about 50 kcal mol⁻¹, trapping the unfolded state⁹. For appreciable refolding to occur, the unfolded domain must shorten to near its native folded length, requiring the force to be reduced to a few piconewtons (see Methods). The average stabilities of immunoglobulin and fibronectin domains are comparable^{15–18}, but variations in stability within both classes will ensure that unfolding is not catastrophic.

At low forces our results are consistent with the model shown in Fig. 1b–d. At high forces yielding is due to domain unfolding, whereas in muscle fibres it is associated with detachment of titin molecules from the thick filament⁷. Perhaps unfolding triggers progressive detachment from the thick filament as a consequence of loss of the tertiary structure presenting the binding site. Does unfolding contribute to extensibility in the physiological range of sarcomere lengths? Muscle fibres exhibit stress-relaxation and hysteresis well below the length at which yielding occurs^{7,14} and this might entail a limited amount of unfolding; it is thus conceivable that the immunoglobulin domains act as shock absorbers.

Note added in proof: Gaub, Bustamante and colleagues have independently reported similar findings^{25,26}. □

Methods

Bead preparation. Antibodies were covalently coupled to aldehyde-poly-styrene latex microspheres, diameter 1.1 μm : 10 μl of the bead suspension ($\sim 7.7 \times 10^{10}$ ml⁻¹) were diluted to 1 ml with 10 mM HEPES buffer, pH 7.5, and 10–50 μl of antibody solution (cell supernatant) was added. After 30 min incubation at room temperature with constant stirring, 10–20 μl BSA solution (10 mg ml⁻¹) in HEPES buffer was added for a further 20 min. The suspension was then sonicated, diluted to 10 ml with 50 mM Tris, 10 mM β -mercaptoetha-

nol solution, pH 7.5, and centrifuged at 10,000g for ~ 10 min. Washing in the Tris buffer was repeated before the beads were suspended in 1 ml 0.3 M KCl, 0.1 mM Na₃, 1 mM EGTA, 0.1 mM DTT, 10 mM histidine, pH 7.5.

Cell-surface preparation. A flow cell (internal volume $\sim 50 \mu\text{l}$) was made from a clean glass coverslip, coated with a thin film of nitrocellulose and a microscope slide, separated by two strip fragments of coverslip (attached by Vaseline). Incubations, each 10 min, were performed at room temperature sequentially as follows: antibody solution (cell supernatant); BSA (10 mg ml⁻¹) in 0.3 M KCl, pH 7.5; titin (0.1 mg ml⁻¹ in 0.3 M KCl, pH 7.5, prepared from rabbit back muscle¹⁹) containing 5 mg ml⁻¹ BSA; bead suspension. The final solution was 0.3 M KCl. In the experiments described the antibodies used bind ~ 100 nm from the M- (AB5; ref. 20) and Z-lines (T12 (ref. 12) or DB1; J.T., unpublished data) *in situ*. Preliminary experiments using a wide variety of antibodies showed that the displacement of beads with a roughly 1-pN force, produced by fluid flow, depended on the distance between epitopes determined *in situ*, thus validating the method.

Density of titin and antibodies on the bead and coverslip. Rough calculations of antibody density on the surfaces, allowing for the likely binding efficiency towards titin, give upper limits for the density of bound titin molecules on the coverslip of 2 per μm^2 , and ~ 8 competent binding sites per bead. The observed density of tethered beads was very much lower than this, about 100–1,000 per mm^2 , and this was not increased by longer incubation times of the beads. Thus, it is likely that most beads were attached by a single linkage.

Equipment. Most of the optical-tweezers technique used has been described previously²². The apparatus was based on a modified inverted microscope (Zeiss Axiovert), using a 2W Nd-YLF laser, 1.047 μm (TFR, SpectraPhysics), with a $\times 63$ Planapo NA 1.4 objective. The slide was mounted on a piezo-electric transducer (pzt)-operated stage (natural frequency ~ 100 Hz) equipped with capacitance gauge transducers. Experimental data were digitized using a PC with an analog-to-digital converter (National, Labview), usually at a sampling frequency of 2 kHz. Trap stiffness was determined from the root-mean-square deviation of the brownian motion of a trapped bead, or from the corner frequency of brownian noise, or the response to a sine-wave input into a stage pzt, with a trapped bead about 5 μm above the coverslip surface.

Dynamic force-extension data. Having trapped a bead and centred it by trial and error, a triangular wave was applied to one pzt, and the quadrant detector and capacitance gauge outputs were recorded. Data were processed as follows: calibrations for the capacitance gauge and the detector were applied and corrections were made for the nonlinearity of the detector and for the nonlinearity of force produced by the trap. Five to sixteen triangular wave sweeps were averaged and filtered at 100 Hz. The four quadrants of data from the triangular wave were averaged and converted into extension and force, assuming that the bead was free to rotate, and correcting for the geometry of the bead and molecule. Analysis programs were written using Matlab.

Static force-extension data. The method used was similar to that described for the dynamic force-extension data, except that length changes were applied using the digital-to-analog output from a PC, giving ramp-faced steps (velocity 250 nm s⁻¹), and the pzt-operated stage was servo-controlled. Beads tended to detach at high maintained force, and the most complete experiments entailed a compromise with the time allowed for the steady state to be established after a stretch (Fig. 3). The bead centring was checked at the end of an experiment, showing that in most cases slippage was negligible.

Polymer models. Least-squares fits to the experimental force-extension data were made for the freely-jointed chain model¹¹ and the worm-like chain model¹² (using the interpolation formula given in ref. 23), allowing all the parameters to vary. A Simplex method (Matlab) was used. The force-extension curve for an unfolded domain was obtained by subtraction of the dynamic curve from the release section of a hysteresis curve, weighted for the number of unfolded domains. The resultant curve was then fitted with a single component freely-jointed chain model to obtain the contour length.

Received 4 March; accepted 14 April 1997.

- Trinick, J. Cytoskeleton: titin as a scaffold and spring. *Curr. Biol.* **6**, 258–260 (1996).
- Maruyama, K. Connectin, an elastic protein of striated muscle. *Biophys. Chem.* **50**, 73–85 (1994).
- Labeit, S. & Kolmerer, B. Titins: giant proteins in charge of muscle ultrastructure and elasticity. *Science* **270**, 293–296 (1995).
- Linke, W. A. *et al.* Towards a molecular understanding of the elasticity of titin. *J. Mol. Biol.* **261**, 62–71 (1996).
- Gautel, M. & Goulding, D. A molecular map of titin/connectin elasticity reveals two different mechanisms acting in series. *FEBS Lett.* **385**, 11–14 (1996).

6. Higuchi, H., Nakauchi, Y., Maruyama, K. & Fujime, S. Characterization of α -connectin from striated muscle by dynamic light scattering. *Biophys. J.* **65**, 1906–1915 (1993).
7. Wang, K., McCarter, R., Wright, J., Beverly, J. & Ramirez-Mitchell, R. Viscoelasticity of the sarcomere matrix of skeletal muscles: the titin-myosin composite filament is a dual-stage molecular spring. *Biophys. J.* **64**, 1161–1177 (1993).
8. Soteriou, A., Clarke, A., Martin, S. & Trinick, J. Titin folding energy and elasticity. *Proc. R. Soc. Lond. B* **254**, 83–86 (1993).
9. Erickson, H. P. Reversible unfolding of fibronectin type-III and immunoglobulin domains provides the structural basis for stretch and elasticity of titin and fibronectin. *Proc. Natl Acad. Sci. USA* **91**, 10114–10118 (1994).
10. Politou, A. S., Thomas, D. J. & Pastore, A. The folding and stability of titin immunoglobulin-like modules, with implications for the mechanism of elasticity. *Biophys. J.* **69**, 2601–2610 (1995).
11. Flory, P. J. in *Statistical Mechanics of Chain Molecules* 316–304 (Hanser, Munich, 1989).
12. Fixman, M. & Kovac, J. Polymer conformational statistics. III. Modified Gaussian models of stiff chains. *J. Chem. Phys.* **56**, 1564–1568 (1973).
13. Kellermayer, M. S. Z. & Granzier, H. L. Elastic properties of single titin molecules made visible through fluorescent F-actin binding. *Biochem. Biophys. Res. Comm.* **221**, 491–497 (1996).
14. Linke, W. A., Bartoo, M. I., Ivenmeyer, M. & Pollack, G. H. Limits of titin extension in single cardiac myofibrils. *J. Muscle Res. Cell Motil.* **17**, 425–438 (1996).
15. Politou, A. S., Gautel, M., Pfuhl, M., Labeit, S. & Pastore, A. Immunoglobulin-type domains of titin: same fold, different stability? *Biochemistry* **33**, 4730–4737 (1994).
16. Fong, S. et al. Structure and stability of an immunoglobulin superfamily domain from twitchin, a muscle protein of the nematode *Caenorhabditis elegans*. *J. Mol. Biol.* **264**, 624–639 (1996).
17. Litvinovich, S. V., Novokhatny, V. V., Brew, S. A. & Ingram, K. C. Reversible unfolding of an isolated heparin and DNA binding fragment, the first type III module from fibronectin. *Biochim. Biophys. Acta* **1119**, 57–62 (1992).
18. Plaxco, K. W., Spitzfaden, C., Campbell, I. D. & Dobson, C. M. Rapid refolding of a proline-rich all-beta-sheet fibronectin type-III module. *Proc. Natl Acad. Sci. USA* **93**, 10703–10706 (1996).
19. Soteriou, A., Gamage, M. & Trinick, J. A survey of the interactions made by the giant protein titin. *J. Cell Sci.* **104**, 119–123 (1993).
20. Whiting, J., Wardale, J. & Trinick, J. Does titin regulate the length of muscle thick filaments. *J. Mol. Biol.* **205**, 263–268 (1989).
21. Fürst, D. O., Osborn, M., Nave, R. & Weber, K. The organisation of titin filaments in the half-sarcomere revealed by monoclonal-antibodies in immunoelectron microscopy: a map of ten nonrepetitive epitopes starting at the Z-line extends close to the M-line. *J. Cell Biol.* **106**, 1563–1572 (1988).
22. Simmons, R. M., Finer, J. T., Chu, S. & Spudis, J. A. Quantitative measurements of force and displacement using an optical trap. *Biophys. J.* **70**, 1813–1822 (1996).
23. Bustamante, C. Entropic elasticity of λ -phage DNA. *Science* **265**, 1599–1600 (1994).
24. Tikhovrebova, L. & Trinick, J. Direct visualization of extensibility in isolated titin molecules. *J. Mol. Biol.* **265**, 100–106 (1997).
25. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M. & Gaub, H. E. Reversible unfolding of individual titin Ig-domains by AFM. *Science* (in the press).
26. Kellermayer, M. S. Z., Smith, S. B., Granzier, H. L. & Bustamante, C. Folding-unfolding transitions in single titin molecules characterized with force-measuring laser tweezers. *Science* (in the press).

Acknowledgements. We thank M. Gautel, W. Gratzer, P. Knight, G. Offer, A. Pastore and D. Thomas for discussions and advice. Supported by BBSRC and NIH (J.T.), MRC (J.A.S. & R.M.S.) and the Human Frontier Science Program (R.M.S.).

Correspondence and requests for materials should be sent to R.M.S. (e-mail: r.simmons@kcl.ac.uk).

Crystal structure of ICAM-2 reveals a distinctive integrin recognition surface

José M. Casasnovas*, Timothy A. Springer*, Jin-huan Liu†‡, Stephen C. Harrison†§ & Jia-huai Wang†‡

* The Center for Blood Research, Harvard Medical School, Department of Pathology, 200 Longwood Avenue, Boston, Massachusetts 02115, USA

† Department of Molecular and Cellular Biology, and § Howard Hughes Medical Institute, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA

Recognition by integrin proteins on the cell surface regulates the adhesive interactions between cells and their surroundings^{1,2}. The structure of the 'I' domain that is found in some but not all integrins, has been determined^{3,4}. However, the only integrin ligands for which structures are known, namely fibronectin and VCAM-1 (refs 5–7), are recognized by integrins that lack I domains. The intercellular adhesion molecules ICAM-1, 2 and 3 are, like VCAM-1, members of the immunoglobulin superfamily (IgSF), but they are recognized by an I domain-containing integrin, lymphocyte-function-associated antigen 1 (LFA-1, or

CD11a/CD18). Here we present the crystal structure of the extracellular region of ICAM-2. The glutamic acid residue at position 37 is critical for LFA-1 binding and is proposed to coordinate the Mg^{2+} ion in the I domain; this Glu 37 is surrounded by a relatively flat recognition surface and lies in a β -strand, whereas the critical aspartic acid residue in VCAM-1 and fibronectin lie in protruding loops. This finding suggests that there are differences in the architecture of recognition sites between integrins that contain or lack I domains. A bend between domains 1 and 2 of ICAM-2 and a tripod-like arrangement of N-linked glycans in the membrane-proximal region of domain 2 may be important for presenting the recognition surface to LFA-1. A model of ICAM-1 based on the ICAM-2 structure provides a framework for understanding its recognition by pathogens.

The extracellular fragment of ICAM-2 containing the two predicted IgSF domains was expressed in lectin-resistant CHO Lec3.2.8.1 cells⁸ to obtain a homogeneous, high-mannose, N-linked glycoform. Crystallization trials were done with native and endoH-treated ICAM-2, and crystals were obtained only with the native protein. The structure was determined by multiple isomorphous replacement (MIR) and refined by XPLOR (see Methods and Table 1).

ICAM-2 has two domains with an immunoglobulin(Ig)-like fold (Fig. 1a). The molecule resembles a hockey stick, with a bend of 35° and a rotation of 152° between the N-terminal, membrane-distal domain 1 (D1), and the C-terminal, membrane-proximal domain 2 (D2), defined by superposition of the Ig framework. With domain 2 oriented perpendicular to the membrane, the edge of the two β -sheets formed by β -strands C and D in domain 1 faces outwards.

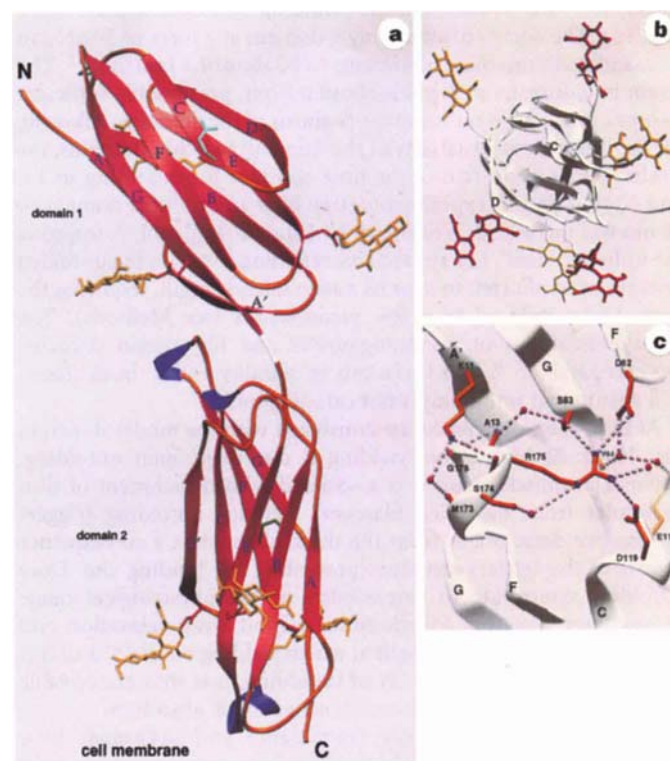


Figure 1 The crystal structure of ICAM-2. **a**, Ribbon diagram with β -strands in red, α -helix in blue and coil in orange. The last residue of domain 1 (Tyr 85) is white. N-linked sugars (yellow), Glu 37 in strand C of domain 1 (light blue) and disulphide bonds (green) are included. **b**, View from the top of domain 1 towards the membrane. Sugars linked to domains 1 and 2 are yellow and red, respectively. **c**, Hydrogen bonds between residues in domain 1 (top) and domain 2 (bottom). Nitrogen and oxygen atoms are in blue and red, respectively. Water molecules are shown as red spheres. Prepared with Ribbons²⁹.

† Present address: Dana-Farber Cancer Institute (J.-h.L. and J.-h.W.) and Children's Hospital (J.-h.W.), Harvard Medical School, 44 Binney Street, Boston, Massachusetts 02115, USA.

This edge bears Glu 37, which is the conserved acidic residue that is important for recognition of ICAMs by LFA-1 (refs 9–12).

ICAM-2 has a compact domain 1 and a relatively larger domain 2 (Fig. 1a). Domain 1 is composed of 85 residues and, like VCAM-1, falls into the I-set category of the IgSF¹³. Unlike typical I-set domains, ICAM2-D1 lacks a short C' strand adjacent to strand C (Fig. 1a). Instead, the C strand turns directly into the CD loop and connects to the D strand across the β -barrel. This part of the molecule is therefore very flat. The four-residue deletion in the CD loop of ICAM2-D1 relative to VCAM-D1 (Fig. 2) presents a striking contrast in the display of the acidic residues that are critical for recognition by integrins. The acidic Glu 37 residue in ICAM-2 is the last residue in β -strand C, whereas the Asp 40 residue in VCAM-1 is at the upturned point of the CD loop (Fig. 2). The remarkably protruding CD loop on which Asp 40 is located in VCAM-1 has no equivalent in ICAM-2. These differences, and the alignment between all β -strands in VCAM-1 and ICAM-2 except for portions of two strands that differ in β -bulges, were correctly predicted on the basis of I-set framework residues (ref. 3, C. Chothia, personal communication). The conformation of the functionally important strands C, D and E in ICAM-2 differs from that in VCAM-1 (Fig. 2). In this part of the structure, the proportion of α carbons within 1.6 Å (asterisked in Fig. 2) is lower than for any other strand in the superimposed domains 1 or 2. There is an 'extra' disulphide bridge between Cys 28 (BC loop) and Cys 71 (end of the F strand) in ICAM-2 that creates a compact tip for domain 1. In VCAM-1, a similar disulphide-induced tapering has been proposed to make the ligand-binding region near the bottom of the domain more accessible⁷.

Domain 2 of ICAM-2 belongs to the C2 set of IgSF domains, which is widely used by cell-surface molecules as building blocks to project the functional N-terminal domain away from the cell membrane¹⁴. As in VCAM-D2 (ref. 15), the conformation of the initial segment of domain 2 is fixed by two main-chain hydrogen bonds between Gln 86 and Glu 116 on the BC loop of domain 2. Glu 116 is followed by a *cis* proline that is conserved in ICAMs and VCAM-1 (Fig. 2). A particularly prominent FG loop in domain 2 of ICAM-2 cradles domain 1 (Fig. 1a). As in VCAM-D2 (refs 6, 7), the C'E and FG loops are unusually long for C2 set domains¹⁴. Except for the C'E and FG loops, ICAM2-D2 and VCAM1-D2 are remarkably similar, in contrast to D1 (Fig. 2).

Six N-linked high-mannose glycans, each predicted to contain about seven carbohydrate residues⁸, are present per ICAM-2 molecule. The first two N-acetyl glucosamine residues are sufficiently ordered to be traced in the electron-density map and included in the structure (Fig. 1). ICAM-2 is among the most heavily glycosylated proteins in terms of sugar/protein ratio to be crystallized so far, with an average of one N-linked site per 32 amino-acid residues. Looking

down the axis of the two IgSF domains, the six N-linked glycans are uniformly distributed around the perimeter of the domains (Fig. 1b). The three glycans in domain 1 are circumferential to the putative binding site around Glu 37 (Fig. 1a). The three glycosylation sites in domain 2 are quite evenly distributed (Fig. 1b, red) around its membrane proximal region (Fig. 1a). This 'tripod' arrangement and the large volume of the complete glycans will create a skirt around the base of D2, which is likely to orient ICAM-2 with respect to the cell surface so that domain 2 is perpendicular to the membrane and the integrin recognition surface in domain 1 is presented to interacting cells. A similar disposition of glycosylation sites is present in the second domain of mouse ICAM-2 (Fig. 2).

Crystallographic analysis of CD2 (ref. 16) and VCAM-1 (refs 6, 7) showed that the two independent copies of the molecules in the crystals differed in the orientation between domains 1 and 2 because of inherent interdomain flexibility. It is difficult to determine how flexible the interdomain linkage in ICAM-2 might be, because there is only one copy of the molecule in the asymmetric unit of the crystal. However, the hydrogen-bond network engaging the A', G and F strands of domain 1 with residues in the FG and BC loops of domain 2 (Fig. 1c) is more extensive than in CD2 or VCAM-1 (refs 6, 7, 16). These interactions and the interdomain hydrophobic contacts involving Ala 13 and Tyr 85 from domain 1, and Leu 170, Leu 172 and Phe 180 from domain 2, may limit interdomain flexibility. In particular, multiple salt-bridges involving more than

Table 1 Crystallographic analysis

	Native	Derivatives		
		C ₂ H ₅ HgCl (0.5 mM)	K ₂ Pt(C ₂ O ₄) ₂ (1 mM)	(CH ₃) ₃ PbAc (25 mM)
Resolution (Å)	15–2.2	15–2.5	15–3.5	15–3.5
Completeness (%)	93.9 (68.8)	90.0	98.0	95.0
Redundancy	3.1	3.0	3.8	2.8
<i>R</i> _{merge} *	0.044 (0.141)	0.055	0.103	0.069
<i>R</i> _{dev} †		0.129	0.207	0.208
PP‡	0.70	0.80	1.0	1.0
<i>R</i> _c §	0.85	0.89		0.70
Refinement				
Model	192 amino-acid residues, 12 NAG and 142 water molecules			
Resolution range (Å)	15.0–2.2			
<i>R</i> factor (%)	22.6 (29.5 free <i>R</i>)			

Completeness and *R* merge for the outermost shell (2.32–2.2 Å) of the native data are in parentheses.

**R*_{merge} = $\sum_i |I_i - \langle I_i \rangle| / \sum_i I_i$.

†*R*_{dev} = $\sum_i |F_o - F_c| / \sum_i F_o$, where *F*_o and *F*_c refer to the measured structure amplitudes of the native and derivatives.

‡PP: phasing power, mean value of the heavy-atom structure amplitudes divided by the residual lack-of-closure.

§*R*_c: Cullis *R* factor for centric reflections; *R*_c = $\sum_i |F_o - F_c| / \sum_i F_o$. Figure of merit was 0.37 for acentric reflections and 0.66 for centric ones.

||Reflections with *F* > 2σ were used in refinement. 10% of reflections were selected for determination of free *R* factor.

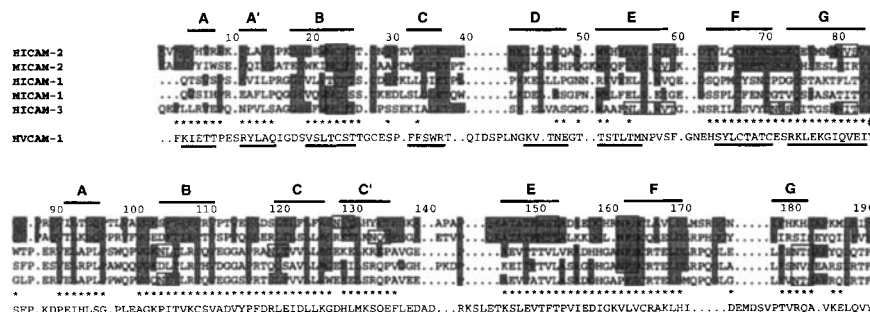


Figure 2 The secondary structure of ICAM-2 and alignment with ICAMs and VCAM-1. β -strands defined by the program DSSP for ICAM-2 and VCAM-1 (ref. 7) are represented by lines. Residues of ICAM-2 conserved in other aligned ICAMs are shaded. Putative N-linked glycosylation sites are framed. ICAM-2 and VCAM-1 were structurally superimposed using the 3DMALIGN module of MODELLER³¹,

independently for domain 1 (top) and domain 2 (bottom). In the structure-based alignment with VCAM-1, only residues with α carbons within 4 Å of one another are aligned, and those within 1.6 Å are marked with asterisks. The conserved Tyr85 (indicated by a hash sign) was close (~1 Å) to its counterpart in VCAM-1 when structural superposition was based on D2.

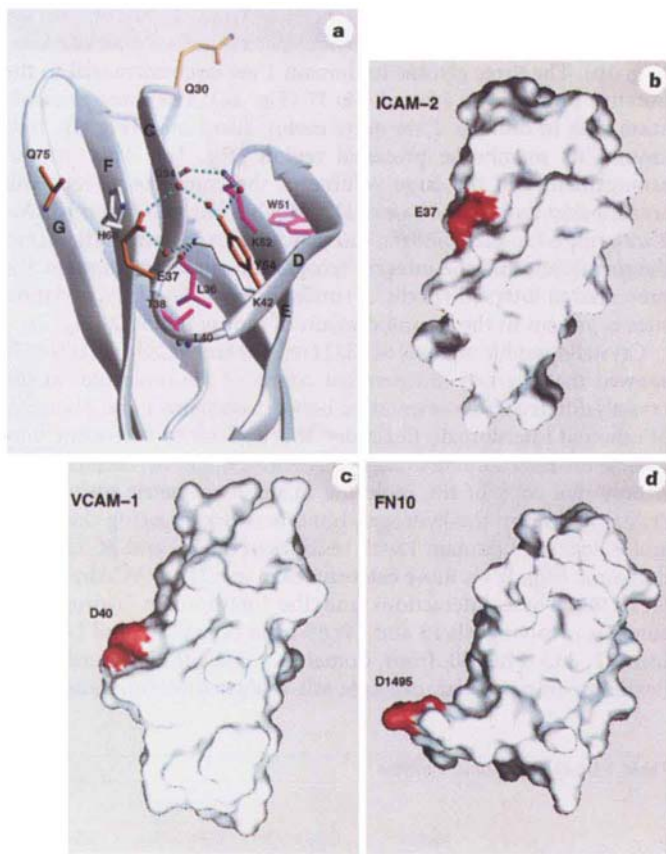


Figure 3 The integrin-binding surface in ICAM-2 and molecular surface representations of integrin ligands. **a**, Residues involved in binding to LFA-1 (orange), domain conformation (magenta) and close to Glu 37 (white) are included. Hydrogen bonds between residues or water molecules (red spheres) are presented as shaded light blue cylinders. Atoms are as in Fig. 1c. Prepared with Ribbons²⁹. **b-d**, Molecular surface representations of integrin-binding domains for ICAM-2, VCAM-1 and fibronectin domain 10 are shown with the acidic residue involved in binding coloured in red. Molecular surfaces were created with GRASP³⁰. Note that the protruding integrin-binding region in VCAM-1 is in the CD loop, whereas in fibronectin it is in the FG loop.

two charged residues¹⁷ centred at Arg 175 (Fig. 1c) are found at the domain interface. The interdomain interactions create a bent conformation for the two-domain module in the crystal structure, with the CD edge of domain 1 and Glu 37 projecting toward the top, optimally presented for adhesive interactions (Fig. 1a). The 35° bend in ICAM-2 compares to bends of 13 to 41° for CD2 (ref. 16) and 5° to 35° for VCAM-1 (refs 6, 7). Also, the rotation between domain 1 and 2 differs, making the C'E loop in domain 2 of ICAM-2 more distant from the CD edge of domain 1 than in VCAM-1.

The strong sequence conservation around Glu 37 in other ICAMs (Fig. 2) suggests that the conformation of their integrin-binding surfaces will be similar to that of ICAM-2. The hydrogen bond between the hydroxyl group of the highly conserved Thr 38 (Fig. 2) and the main chain -NH of Leu 40 positions Glu 37 and helps to determine the CD loop conformation (Fig. 3a). Mutagenesis studies of ICAMs⁹⁻¹² have shown that four conserved residues are important for adhesive interactions with LFA-1 (orange in Fig. 3a). Glu 37, the most critical residue for integrin binding, is the last residue of the C strand in domain 1 of ICAM-2. Tyr 54 on the E strand and Gln 30 in the BC loop are located at the exposed CD edge of the β -barrel. Gln 75 is on the G strand, with its side chain facing Glu 37. These residues define a broad area that includes the CD edge and the

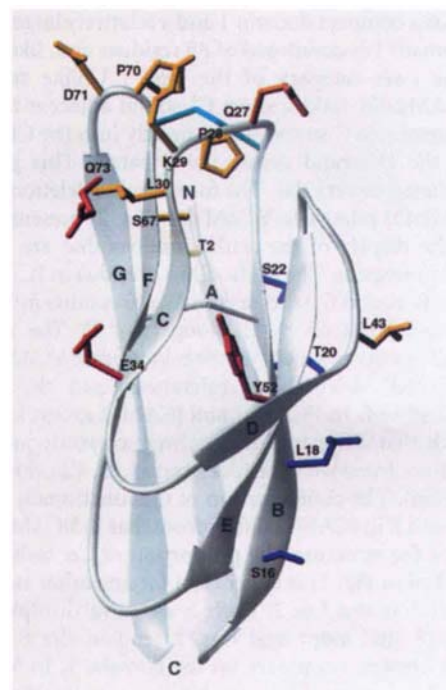


Figure 4 Ligand recognition by ICAM-1. A model of the N-terminal domain of ICAM-1 was constructed with a segment modelling program (Look, Molecular Applications Group) based on the alignment with ICAM-2 shown in Fig. 2. Residues involved in rhinovirus, LFA-1 and PFIE binding are coloured in yellow, red and blue, respectively. Q27, a residue important for LFA-1 and HRV binding, is in orange. Single amino-acid mutations, except the double mutation T20 and S22, with a significant influence in binding (>40% decrease binding compared with wild type) have been included. The 'extra' disulphide bond in the domain is in cyan.

GFC- β -sheet of the domain. Structural and functional data show that residues in the I domain of LFA-1 involved in binding to ICAM-1 lie on both sides of the bound Mg^{2+} (refs 3, 4). By contrast, the disposition of the binding residues in ICAMs and the I domain of LFA-1 are consistent with the notion that Glu 37 in ICAM-2 and the Mg^{2+} in the I domain could coordinate in the centre of the binding interface.

The flat surface of ICAM-2 surrounding Glu 37 (Fig. 3b) complements well in shape the relatively flat surface surrounding the Mg^{2+} in the I domains of Mac-1 and LFA-1 (refs 3, 4). By contrast, in the non-I-domain integrin ligands VCAM-1 (Fig. 3c) and fibronectin (Fig. 3d), the Asp residues involved in binding are located in protruding loops⁵⁻⁷. These differences in presentation of integrin-binding motifs by integrin ligands may therefore reflect general differences in the ligand-binding site between I-domain- and non-I-domain-containing integrins. This difference correlates with the prediction that non-I-domain integrin ligands bind to a β -propeller domain in integrin α -subunits, which contains a putative binding pocket¹⁹.

In addition to serving as a ligand for LFA-1, ICAM-1 is the cell-surface receptor for the major group of human rhinovirus (HRV)^{20,21} and a receptor for sequestration of *Plasmodium falciparum*-infected

erythrocytes in the peripheral vasculature²². Residues involved in the recognition of all these three ligands lie mainly in the N-terminal domain of the receptor^{9,23–26}. A model of ICAM-1 based on domain 1 of ICAM-2 suggests that the rhinovirus^{9,23,24} and LFA-1 (refs 9, 11) binding surfaces are distinct but overlapping (yellow and red, respectively, in Fig. 4), as previously predicted from mutagenesis data. By contrast, residues involved in binding to *P. falciparum*-infected erythrocytes (blue)^{25,26} are in the B strand, on the edge of the domain opposite the LFA-1 binding surface. *Plasmodium*-derived proteins seem to recognize a linear and narrow epitope surface, consistent with the ability of peptides like G15SVLV to block ICAM-1 binding²⁵. Lack of glycosylation in domain 1 of human ICAM-1 compared to ICAM-2 and ICAM-3 may be an important determinant of the promiscuity of this cell-adhesion molecule in terms of recognition by microorganisms. The crystal structure of ICAM-2 may therefore advance the development of therapeutics for treatment of immunological and vascular disorders and pathogenic infections. □

Methods

Protein expression, purification and crystallization. A recombinant ICAM-2 cDNA with a translation stop codon in the position of Glu 196 was generated by PCR from the plasmid pCDIC2.27 (ref. 27), and introduced into the unique *Xho*I–*Not*I site of pBJ5-GS²⁸, to obtain pBJ5-GS/ICAM-2. A CHO cell mutant (CHO Lec3.2.8.1) with highly restricted glycosylation was transfected with pBJ5-GS/ICAM-2 using calcium phosphate, and clones secreting soluble ICAM-2 were selected²⁸. ICAM-2 was purified from the cell supernatant by immunoaffinity with CBR-IC2/2 mAb and size-exclusion chromatography, essentially as described for ICAM-1 (ref. 28). Two major species of 37.5K and 35.9K were seen in SDS–PAGE and were cleaved to 27.8K and 26.4K with endoglycosidase H and N-glycanase, respectively. The N-terminal sequence of ICAM-2, which had not been previously determined, is KVFEV. Therefore, the signal sequence is three residues longer than previously predicted from the cDNA sequence²⁷ and the recombinant soluble ICAM-2 contains 192 amino acid residues (Fig. 2).

Crystals were grown best at 4°C using the hanging-drop vapour diffusion method, with a protein solution of 15–17 mg ml⁻¹ and a crystallization solution with 20% PEG 4000, 20 mM cacodylate buffer, pH 6.4, and 25 mM β -octyl-glucopyranoside. They belong to space group C2 ($a = 69.0$ Å, $b = 63.7$ Å, $c = 76.2$ Å and $\beta = 114.5^\circ$), with one molecule per asymmetric unit and 45% solvent by volume. Crystals were frozen under -160°C liquid nitrogen cold stream for data collection using 20% ethylene glycol as cryoprotectant.

Structure determination and refinement. Each data set was collected from a single frozen crystal, using a Siemens/Nicolet area detector equipped with an Elliott GX-13 rotating anode X-ray generator. Raw data were reduced using XDS and further processed with CCP4 suite (SERC collaborative computer project 4; Daresbury Laboratory, Daresbury, UK).

The structure was determined by the MIR method using three heavy-atom derivatives: ethyl mercury chloride, trimethyl lead acetate and potassium bis(oxalato)-platinate. Heavy-atom parameters were determined by difference Patterson and difference Fourier and refined with HEAVY and MLPHARE. Phase statistics are listed in Table 1. The initial MIR map at 3.5 Å was of poor quality, although protein–solvent boundaries were detected. This map was dramatically improved by applying a density modification procedure as implemented in the CCP4 program DM (SERC CCP4). After the first model was built, a second run of DM with a more accurate envelope further improved the map on which a complete ICAM-2 molecule was built, except for the C'E-loop of D2.

X-PLOR (a system for crystallography and NMR, New Haven, CT, version 3.1) was used to carry out the structure refinement, alternated by manual model rebuilding. A resolution range from 15 to 3 Å was first used and was extended stepwise to 2.2 Å. Both energy minimization and simulated annealing options were applied and at a very late stage only energy minimization was done. Solvent mask correction was made during refinement. An overall anisotropic temperature factor was applied to F_o to obtain the best results.

The current model contains all 192 amino-acid residues, 12 N-acetyl glucosamines (NAG) residues and 142 water molecules with good geometry (r.m.s. deviations for bond length and bond angles are 0.019 Å and 2.2°, respectively). Only two residues (Ala 140 and Gln 145) that lie on the poorly defined C'E loop have disallowed main-chain torsion angles in the Ramachandran plot of our model. Quite good density is present for the first NAG at high contour (1.2 σ) and for the second one at low contour (>0.7 σ). The missing sugar residues in our model account for 20% of the molecular weight of the protein and may contribute to the R factor of our model, particularly at the low-resolution bin.

Received 7 January; accepted 17 March 1997.

- Hynes, R. O. Integrins: Versatility, modulation, and signaling in cell adhesion. *Cell* **69**, 11–25 (1992).
- Springer, T. A. Traffic signals for lymphocyte recirculation and leukocyte emigration: The multi-step paradigm. *Cell* **76**, 301–314 (1994).
- Lee, J.-O., Rieu, P., Arnaout, M. A. & Liddington, R. Crystal structure of the A domain from the α -subunit of integrin CR3 (CD11b/CD18). *Cell* **80**, 631–638 (1995).
- Qu, A. & Leahy, D. J. Crystal structure of the I-domain from the CD11a/CD18 (LFA-1, α 2) integrin. *Proc. Natl Acad. Sci. USA* **92**, 10277–10281 (1995).
- Leahy, D. J., Aukhil, I. & Erickson, H. P. 2.0 Å crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region. *Cell* **84**, 155–164 (1996).
- Jones, E. Y. *et al.* Crystal structure of an integrin-binding fragment of vascular cell adhesion molecule-1 at 1.8 Å resolution. *Nature* **373**, 539–544 (1995).
- Wang, J.-h. *et al.* The crystal structure of an N-terminal two domain fragment of VCAM-1: A cyclic peptide based on the domain 1 C-D loop can inhibit VCAM-1/ α 4 integrin. *Proc. Natl Acad. Sci. USA* **92**, 5714–5718 (1995).
- Stanley, P. Chinese hamster ovary cell mutants with multiple glycosylation defects for production of glycoproteins with minimal carbohydrate heterogeneity. *Mol. Cell. Biol.* **9**, 377–383 (1989).
- Staunton, D. E., Dustin, M. L., Erickson, H. P. & Springer, T. A. The arrangement of the immunoglobulin-like domains of ICAM-1 and the binding sites for LFA-1 and rhinovirus. *Cell* **61**, 243–254 (1990).
- Holness, C. L. *et al.* Analysis of the binding site on intercellular adhesion molecule 3 for the leukocyte integrin lymphocyte function-associated antigen 1. *J. Biol. Chem.* **270**, 877–884 (1995).
- Klickstein, L. B., York, M. B., de Fougerolles, A. R. & Springer, T. A. Localization of the binding site on intercellular adhesion molecule-3 (ICAM-3) for lymphocyte function-associated antigen-1 (LFA-1). *J. Biol. Chem.* **271**, 23920–23927 (1996).
- Sadhu, C. *et al.* LFA-1 binding site in ICAM-3 contains a conserved motif and non-contiguous amino acids. *Cell Adhes. Commun.* **2**, 429–440 (1994).
- Harpaz, Y. & Chothia, C. Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains. *J. Mol. Biol.* **238**, 528–539 (1994).
- Williams, A. F. A year in the life of the immunoglobulin superfamily. *Immunol. Today* **8**, 298–303 (1987).
- Wang, J.-h. *et al.* Structure of a functional fragment of VCAM-1 refined at 1.9 Å resolution. *Acta Crystallogr.* **52**, 369–379 (1996).
- Bodian, D. L., Jones, E. Y., Harlos, K., Stuart, D. I. & Davis, S. J. Crystal structure of the extracellular domain of the human cell adhesion molecule CD2 at 2.5 Å resolution. *Structure* **2**, 755–766 (1995).
- Musafia, B., Buchner, V. & Arad, D. Complex salt bridges in proteins: Statistical analysis of structure and function. *J. Mol. Biol.* **254**, 761–770 (1995).
- Huang, C. & Springer, T. A. A binding interface on the I domain of lymphocyte function associated antigen-1 (LFA-1) required for specific interaction with intercellular adhesion molecule 1 (ICAM-1). *J. Biol. Chem.* **270**, 19008–19016 (1995).
- Springer, T. A. Folding of the N-terminal, ligand-binding region of integrin α -subunits into a β -propeller domain. *Proc. Natl Acad. Sci. USA* **94**, 65–72 (1997).
- Greve, J. M. *et al.* The major human rhinovirus receptor is ICAM-1. *Cell* **56**, 839–847 (1989).
- Staunton, D. E. *et al.* A cell adhesion molecule, ICAM-1, is the major surface receptor for rhinoviruses. *Cell* **56**, 849–853 (1989).
- Berendt, A. R., Simmons, D. L., Tansey, J., Newbold, C. I. & Marsh, K. Intercellular adhesion molecule-1 is an endothelial cell adhesion receptor for *Plasmodium falciparum*. *Nature* **341**, 57–59 (1989).
- McClelland, A. *et al.* Identification of monoclonal antibody epitopes and critical residues for rhinovirus binding in domain 1 of intercellular adhesion molecule 1. *Proc. Natl Acad. Sci. USA* **88**, 7993–7997 (1991).
- Register, R. B., Uncapher, C. R., Naylor, A. M., Lineberger, D. W. & Colonno, R. J. Human-murine chimeras of ICAM-1 identify amino acid residues critical for rhinovirus and antibody binding. *J. Virol.* **65**, 6589–6596 (1991).
- Ockenhouse, C. F., Betageri, R., Springer, T. A. & Staunton, D. E. *Plasmodium falciparum*-infected erythrocytes bind ICAM-1 at a site distinct from LFA-1, Mac-1, and human rhinovirus. *Cell* **68**, 63–69 (1992).
- Berendt, A. R. *et al.* The binding site on ICAM-1 for *Plasmodium falciparum*-infected erythrocytes overlaps, but is distinct from, the LFA-1-binding site. *Cell* **68**, 71–81 (1992).
- Staunton, D. E., Dustin, M. L. & Springer, T. A. Functional cloning of ICAM-2, a cell adhesion ligand for LFA-1 homologous to ICAM-1. *Nature* **339**, 61–64 (1989).
- Casasnovas, J. M. & Springer, T. A. Kinetics and thermodynamics of virus binding to receptor: Studies with rhinovirus, intercellular adhesion molecule-1 (ICAM-1), and surface plasmon resonance. *J. Biol. Chem.* **270**, 13216–13224 (1995).
- Carson, M. Ribbon models of macromolecules. *J. Mol. Graph.* **5**, 103–106 (1987).
- Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).
- Sali, A. Thesis, Univ. London, 1991.

Acknowledgements. We thank S. Martin and D. Staunton for recombinant cDNA, J. Bickford for technical assistance; P. Stanley for mutant CHO cells; and the members of S.C.H.'s group for helpful discussions, especially T. Stehle, D. Rodgers and B. Harris. This work was supported by grants from the NIH. S.C.H. is an investigator of the Howard Hughes Medical Institute.

Correspondence should be addressed to J.H.W. and requests for materials should be addressed to T.A.S. The ID code for the ICAM-2 coordinates in the Protein Data Bank is 1ZXQ.

Notes for contributors

Nature is a weekly international journal covering all the sciences. Its readership is interdisciplinary, so manuscripts should be written clearly and simply, authors remembering that English is not the first language of many readers. Competition for *Nature's* limited space is severe, so brevity is essential. *Nature's* main formats for original research are Articles and Letters to *Nature*. Lengths of these contributions are specified in terms of numbers of pages they take up. An uninterrupted page of text contains about 1,300 words, so a typical Letter to *Nature*, occupying 2 1/2 pages, can contain about 1,500 words of text and four modest-sized display items (figures and/or tables) with brief legends. There are no formal limits on the number of display items, but extra items must be accompanied by commensurate reductions in the length of the text. Length limits can, however, be varied at the editors' discretion.

Articles are reports of original scientific research whose conclusions represent a substantial advance in understanding of an important problem and are of broad general interest. They do not generally exceed 5 pages of *Nature* and have no more than 50 references. They have a 'heading' of up to 80 words, often rewritten by *Nature's* editors, which advertises the content of the paper in general terms and which should not contain numbers, abbreviations or measurements unless central to the message of the paper. The first three or four paragraphs of the main text should introduce and summarize the findings. Articles have a few subheadings of not more than 45 characters each.

Letters to *Nature* are short reports of original research focused on an outstanding finding with implications of interest to nonspecialists. They should not generally exceed 2 1/2 pages of *Nature*. They have a bold first paragraph of fewer than 180 words summarizing the background, rationale, conclusion and main implication of the work, aimed particularly at nonspecialist readers. Letters should have no more than 30 references.

Preparation of Articles and Letters

All manuscripts should be in English, typed double-spaced on one side of the paper only. Manuscripts (and revised manuscripts) should not be submitted in electronic form.

Manuscripts should be accompanied by two copies of a brief covering letter from the corresponding author, containing a full postal address, telephone and fax numbers, and e-mail address. Articles and Letters contain a statement at the end of the text: "Correspondence and requests for materials to xxxxx", with an e-mail address if desired, followed by the accession numbers or Web site address for supporting datasets. The designated author need not be the senior author, but *Nature* will expect this author to handle matters arising from the published contribution.

Five copies of manuscripts (and revised manuscripts) and figures are required. Each copy of a revised manuscript should be accompanied by a copy of the author's response to referees' and editors' comments on the earlier version. Two copies of relevant papers by any of the authors submitted or in press elsewhere (including papers submitted elsewhere while the *Nature* manuscript is under consideration) should also be enclosed, clearly marked as such; failure to do so may prejudice consideration of the manuscript.

Titles should not exceed three lines of 28 characters each (including spaces), and should avoid numerical values, acronyms, abbreviations and punctuation wherever possible.

References should be numbered sequentially as they appear in the text, tables and figure legends. Titles of all cited articles should be provided. Only papers that have been published or submitted to a named journal should be in the reference list: manuscripts in preparation and personal communications should be mentioned in the text with a list of authors. Published conference abstracts may be included in the reference list, but text, footnotes, grant details and acknowledgements may not.

Acknowledgements must be brief and appear after the reference list. *Nature* does not publish grant or contribution numbers.

Tables should each be presented on a separate sheet of paper. Symbols and abbreviations should be defined immediately below the table, fol-

lowed by essential descriptive material, all in double-spaced text.

Figures should not be larger than 22 by 28 cm unless unavoidable, and should not contain more than one panel unless the assembled parts are logically connected. Each copy should be marked with the corresponding author's name and, when known, the manuscript reference number. One photocopy of original figures should be provided. All original figures are returned when *Nature* cannot offer to publish a manuscript, but one copy of the manuscript and the photocopy of the figures are kept in confidential files. Colour figures are welcome, but a contribution towards the cost of reproduction is requested. Inability to pay this charge will not prevent *Nature* from publishing figures for which colour is essential. Once a paper is accepted for publication, electronic versions of figures are preferred.

Methods. Brief details of materials and methods can be given in the text. Otherwise, they should be described at the end of the text in a section entitled "Methods", subdivided by short, bold headings referring to the methods used. This section should not exceed 800 words and ideally should be shorter. In contributions with methods sections, figure legends should not exceed 100 words each, and should be confined to a description of the figure and the symbols used.

Materials. As a condition of publication, authors are required to make materials and methods used freely available to academic researchers for their own use. Supporting datasets must be made available by the authors both directly and by posting on *Nature's* Web site, by deposition in the appropriate public database or by distribution on the Internet.

Submissions. Manuscripts can be submitted to the Editor at *Nature*, Porters South, 4 Crinan St, London N1 9XW or at *Nature*, 968 National Press Building, Washington, DC 20045-1938. Proofs should be returned by express mail to London. Fax numbers and e-mail addresses for all editorial offices are on *Nature's* information page.

Presubmission enquiries for Articles and Letters should be made by e-mail to nature@nature.com, clearly labelled as such. They should include a fully referenced summary paragraph in *Nature* format, and be accompanied by a paragraph stating briefly why the submission will be of interest to a broad general scientific readership. Authors will usually receive a response within two working days of receipt.

Material submitted to *Nature* must not be discussed with the media. Accepted contributions may be discussed with the media a week before the publication date under *Nature's* embargo conditions. (See *Nature* 383, 745; 1996 for further explanation.)

Other contributions

Nature publishes other contributed material as follows:

Correspondence contributions are short comments on topical issues, anecdotal material or reactions of readers to non-technical material published in *Nature*. All are unsolicited; submissions by e-mail to corres@nature.com are preferred (accents should be clearly indicated).

Commentary articles are informal editorials on topical issues of public interest that have some bearing on scientific research. They are often controversial. Unsolicited manuscripts are considered, but proposals can be made to the Commentary Editor by e-mail (nature@nature.com).

News and Views articles inform nonspecialist readers about new scientific advances, as reported either in published papers or at scientific meetings. Most are commissioned, but proposals can be made to the News and Views Editor before the paper is published or the meeting takes place. Detailed guidelines are available on request.

Scientific Correspondence contributions are brief reports of general interest, and are usually peer-reviewed. Priority is given to contributions of fewer than 500 words. Detailed guidelines are available on request. Presubmission enquiries are not considered: contributions for Scientific Correspondence should be submitted by mail, not e-mail.

Review Articles survey recent developments in a topical area of research. They should not generally occupy more than 6 pages of *Nature*.

Progress articles are shorter reviews of particularly topical and fast-moving fields. They do not exceed 4 pages of *Nature*.

Contributors wishing to submit a Review Article or Progress should first send an enquiry or a brief synopsis to the Reviews Coordinator, from whom guidelines are available on request.

In and around the lab

This week's pot-pourri for the laboratory includes a freeze dryer, a centrifugal concentrator, a chemical reaction system, a radio transmitting data logger and a miniature spectrophotometer.

Lab Renovation Guide

From Fisher Hamilton

A 12-page brochure to help laboratory managers develop awareness of the many variables involved in renovation projects

Awareness of all the factors involved in designing and managing a lab can contribute to realistic expectations, efficient planning, smart purchasing and, ultimately, a safe and efficient laboratory. The guide defines the process in four phases: conceptual, planning, procurement and installation. The conceptual phase lays the groundwork, and includes suggestions on developing a master plan, assembling a renovation team, budget and cost control, scheduling and safety. A helpful list of pre-design, considerations, such as usage patterns and American Disabilities Act requirements are also included. Checklists help readers work through the planning phase, where they are asked to consider planned utilities, safety equipment, instrumentation, fume hoods, casework, work surfaces, sinks, fixtures and accessories. A typical 12-month renovation schedule is provided as a visual reference to project planning. An overview of considerations and activities associated with procurement and installation phases rounds out the piece. A reference chart with pertinent dimensions also is included.

Reader Enquiry No. 100

UVlink CL 508 crosslinker

From UVItec

This crosslinker is a compact unit occupying minimum bench space with a footprint of only 350 mm × 350 mm and a usable interior chamber of 270 mm × 300 mm

Designed with microprocessor control, this unit is intended for binding nucleic acids to membranes. The correct ultraviolet dosage can be set using the membrane switch keypad, in either energy units (Joules) or time (seconds). There are nine pre-set settings for energy exposure and nine for time exposure, as well as operator selection for



UVItec's ultraviolet crosslinking unit.

manual control in either units. The ultraviolet energy is continuously monitored by a microprocessor-controlled, photo-feedback system, which compensates for variation in output from the ultraviolet sources, particularly tube ageing. In this way, consistency of operation and maximum efficiency is maintained. The CL 508 door is safety interlocked against opening during operation and the observation window in the door blocks ultraviolet. Time savings over the conventional vacuum oven baking method are said to be considerable (seconds or minutes, as opposed to hours). Applications include fixing of nucleic acids to nylon or nitrocellulose membranes, Southern or northern blotting or dot blotting and colony or plaque lifts, elimination or reduction of PCR contamination, nicking ethidium bromide stained DNA in agarose gels, gene mapping for creating cleavage inhibiting thymine dimers, screening RecA mutations, ultraviolet curing of polymers, adhesives and inks, and ultraviolet sterilization.

Reader Enquiry No. 101

Virtual Filing Cabinet

From Institute of Physics

A redesigned web site that features a virtual filing cabinet enables physicists to annotate, label and store articles taken from any of 31 research journals

New features on this site include menu personalization, preset search facilities, the means to configure bespoke PostScript downloads and a facility to create a table of contents e-mail alerting service. The redesigned site (<http://www.iop.org>) has been further enhanced by improved navigation which helps to provide swifter retrieval of information.

Reader Enquiry No. 102

VersaBlot

From Asys Hitech

An automated western blot processor equipped with five or six high precision peristaltic pumps

Dispensing different reagents, up to five or six, with different volumes, time of incubation, rocking and aspiration can be preprogrammed by a large keyboard and an alphanumeric display, corresponding to user needs. In detection of infectious diseases safety is one of the most important aspects for the lab personnel. If the safety lid is opened during operation VersaBlot stops immediately and will



VersaBlot for western blot processing protocols.

remain inoperative until the cover is closed. Acoustic alarms inform the operator if the waste bottle is full or reagent bottles have to be refilled. Aspiration is performed by a high-speed peristaltic pump. VersaBlot 'speaks' English, French, Spanish, Italian and German.

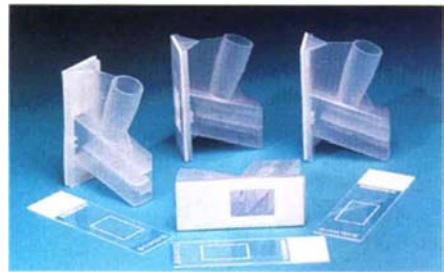
Reader Enquiry No. 103

Megafunnel

From Shandon

The new funnel provides a larger cell deposition area on a microscope slide

Reducing the need to pre-spin large volumes of fluid, this funnel is particularly suitable for urines and larger volume specimens. Each Megafunnel pack comes complete with specially coated Cytoslides, which are designed to improve cellular adhesion. Funnel and slide are sealed with a foam gasket, allowing fluid to be retained and subsequently decanted from the funnel after centrifugation. Since this avoids drying of the specimen during preparation, the resulting specimen is suitable for Papanicolaou staining, if required. Cells from 6 ml of fluid can be deposited on the slide in a single spin. Special baffles have been strategically placed inside the Megafunnel to ensure the cells remain evenly distributed. To aid even distribution further, additional absorbent filter material has been placed at each of the corners around the rectangular



Megafunnel for cell deposition on slides.

opening. The introduction of the new Megafunnel means that Cytospin users can now choose the optimum method for the preparation of their specimens.

Reader Enquiry No. 104

Unitop freeze dryer

From Virtis

The unit benefits from new chamber geometry, which allows for more efficient use of the available chamber volume

In practical terms, the change from a cylindrical drying chamber to square allows the use of additional shelf space for many applications, thereby increasing potential batch sizes. Available through Biopharma Process System, the Unitop is the entry-level shelf freeze dryer, with a maximum shelf area from 0.5 m² for stoppering applications and 0.78 m² for bulk applications. This represents five shelves for stoppering and six shelves for bulk applications. The shelf respacing option allows the purchase of a hydraulic stoppering freeze dryer with five shelves for small vials, but with the ability to store shelves at the bottom of the chamber, which enables the use of larger vials on fewer shelves. The Unitop shelves can be thermoregulated from -45 °C to +65 °C, whilst using 'green' refrigerants. Accuracy is also unaffected at ± 0.5 °C across the whole shelf stack.

Reader Enquiry No. 105

General chemistry

PHM290 pH-stat controller

Radiometer

pH-stat controller also performs titration

This controller has been designed for versatility, providing reliable stat and end-point titrations, as well as accurate pH and mV measurements. Using this system for stat titrations, the set-point value is said to be obtained quickly, without overshoot, and can be maintained within a few mV for several hours. This makes the PHM290 suitable for chemical processes that require a specific pH environment. The system maintains the pH in enzyme activity studies and



Radiometer's pH-Stat also performs titrations.

neutralization processes. For accurate titrant addition, the PHM290 can be connected to an autoburette, which is said to provide high resolution with microlitre accuracy. The speed at which the reagent is added is controlled by one of two special modes. The adaptive addition algorithm is a new mode dedicated to samples with a higher buffer capacity, whereas the proportional integral derivative is suitable for weakly buffered samples. A magnetic valve can be connected for larger experimental setups that require greater rates of addition. A full report can be generated, including the pH, volume of titrant added, rate of addition, temperature and date. This can be printed out, or downloaded to a PC at user-defined time intervals, with the last 600 curve points stored in the memory. Furthermore, a macro for use with Lotus 1-2-3 or Microsoft Excel is supplied. Called Statsheet PC software, it displays 'live curves' of pH/mV versus time, as well as volume of titrant versus time.

Reader Enquiry No. 106

UniVapo 150 centrifugal concentrator

From Integrated Separation Systems

For high-efficiency concentration and evaporation of biological fluids

This system combines vacuum and centrifugal force to provide a gentle, yet effective, method for the concentration and evaporation of biological fluids. It offers a swing-out rotor that holds microplates, as well as a wide selection of other rotor types. Moreover, it is designed and manufactured to be completely air-tight, with a cast aluminum vacuum chamber and a thick, clear plastic lid. Centrifuge surfaces are treated with oven-baked, corrosion-resistant powder paint to prevent problems associated with using aggressive solvents or solutions. The centrifuge uses a strong eight-field magnetic drive to provide fast acceleration and brings the rotor to a quick, smooth stop, typically within nine seconds.

Reader Enquiry No. 107

LabMax

From Mettler Toledo

An automated lab reactor system for chemical syntheses

Mettler Toledo has recently introduced a new system designed to automate many of the functions of chemical syntheses. This turn-key system handles a variety of routine, repetitive tasks, including data logging, temperature control, agitation speed, liquid, gaseous and solid feed rates, distillation and reflux and pH control. A selection of jacket reaction vessels is available to handle different reaction volumes and experimental conditions. Utilizing both Camile TG and Microsoft Windows95 software, a simple or complex experimental recipe can easily be



LabMax, a chemical synthesis system.

implemented. The resulting computer-controlled system allows precise control and repetition of sensitive processes, such as exothermic reactions, crystallization, pH or temperature-dependent additions, thus providing extremely high reproducibility from experiment to experiment. Complete data logging occurs throughout experiments assuring that no critical or unexpected action is missed. The reactor will operate completely unattended, thereby increasing operator productivity. It also relieves the user of day-to-day chemical manipulations, freeing them for other, perhaps more creative, tasks.

Reader Enquiry No. 108

Interface/MS LGC

From LGC

Support and accessories for the chemical reaction interface/MS (CRIMS)

CRIMS is an analytical method that can allow the replacement of radioactive isotopes in drug metabolism studies. The system consists of a universal interface for the formation of a dry particle beam, a one- or two-stage momentum separator, a CRIMS interface and a microwave power supply. The chemical reaction interface involves the combustion and atomization of analyses within a microwave-induced plasma oven. These atoms then combine with a reactant gas to form small, stable oxidation products that flow into a mass spectrometer. This array of low molecular weight gaseous species is detected by the mass spectrometer using selected ion monitoring. For example, m/z 31 is used to monitor ^{15}NO and m/z 45 for $^{13}\text{CO}_2$. CRIMS is said to have been used successfully to study metabolism in human growth hormone.

Reader Enquiry No. 109

Data collection

Icespy RL

From Labdata

A new radio-transmitting data logger

Installation of conventional data loggers with hard-wired probes can be time consuming and inconvenient. The new Icespy RL is a data logger that transmits its infor-



A radio-transmitting data logger from Labdata.

mation by a radio signal to be received and stored by a collection box, therefore requiring minimal installation. This economically priced unit will record 20,000 readings from transmitter units in up to ten locations with temperatures ranging from -40°C to $+135^{\circ}\text{C}$. The data can then be downloaded using a Windows software package, which will automatically produce graphs of the results and has the option of exporting to a spreadsheet package. Online monitoring is possible enabling viewing of temperatures on your computer as the Icespy is recording.

Reader Enquiry No. 110

Labcold Monitor Major 2000

From Borolabs

The new monitor can now log up to 192 channels of information from laboratory equipment

The Major 2000 is a recording system capable of monitoring, analysing and presenting data up to the standards required for the US Food and Drug Administration or good laboratory practice. Its main function is to provide continuous temperature information from freezers, refrigerators and incubators and to alert the technician should the temperature vary from pre-set limits. Each of the 192-channel temperatures can be displayed digitally on the monitor to a resolution of $\pm 0.1^{\circ}\text{C}$. Should the temperature go outside the range individually set by the user, the monitor registers an alarm state. An adjustable time delay has also been incorporated to allow the user to avoid nuisance alarms, such as during defrosting. As an alternative to the standard temperature sensor, it is also possible to monitor any device where an on/off signal exists. An equipment probe can be installed to monitor such functions as pH and CO_2 levels. For added versatility the monitor can support a radio pager to report a fault. Another advantage of the increase in the number of channels is the fact that the monitor can be shared between departments to reduce cost.

Reader Enquiry No. 111

ADC64 PCI bus card

From Adept Scientific

Simultaneous sampling DSP data acquisition on PCI bus

A digital signal processing (DSP) based PCI bus data acquisition card is the latest addition to the Innovative Integration range, which is supplied and supported in the UK by Adept Scientific. The new ADC64 is used for PC-continuous, multi-channel 16-bit data acquisition, bringing together a high-performance, 32-bit floating point DSP core, a very broad array of analog and digital I/O, multiple A/D converters and bus-mastering PCI interface. Another feature of the card is its use of multiple A/D converters. This avoids the phase errors associated with single A/D converter boards that lead to misinterpretation of data. The ADC64 performs conversions on up to eight channels simultaneously. Up to 64 16-bit, instrumentation-grade analog input channels are available, with 120 dB/decade antialias filters and programmable gain, capable of sampling at any rate up to 100,000 samples per sec. The board also provides two channels of 16-bit, instrumentation-grade analog output running at up to 100 kHz per channel. Sixteen bits of high-drive digital I/O are also on-board. Fast, continuous, gap-free multichannel data acquisition under Microsoft Windows is achieved using the ADC64 with DASyLab software. DASyLab is a multifunctional program that makes it easy to create anything from simple data logging and display applications through to complex, high-speed acquisition and frequency domain analysis, and even mimic diagrams of complete test rigs or processes.

Reader Enquiry No. 112

Working with tissue

EMS 4000

From Electron Microscopy Sciences

An automatic oscillating tissue slicer with built-in refrigeration

The EMS 4000 is designed for sectioning either fixed or fresh tissue, as thin as ten micrometres. It is said to eliminate the need for embedding or freezing samples, thus reducing preparation time. The risks of distortion and artefacts, normally associated with these procedures are also eliminated. For maximum stability, an ultralight blade mechanism mounted on a heavily reinforced chassis has been engineered, which is said to eliminate vibration. For precision, this tissue slicer features thumbwheel selection of section thickness, as well as automatic control circuitry so that consecutive sections may be cut easily and rapidly. In addition, the unit features motorized control of the blade height and a complete digi-

tal display of the total cutting excursion. The blade speed is precise with a variety of adjustable speeds and the specimen advance is motorized with a variable speed option as well. The refrigeration unit is built into the EMS 4000 cabinet and maintains the tray temperature near 0°C during the sectioning process and offers maximum tissue life and best cutting consistency.

Reader Enquiry No. 113

Digitimer Multistim

From Medical Systems

A multichannel tissue-bath stimulator system

The Multistim has been designed to facilitate multiple isolated tissue-bath stimulation, with flexible control of stimulation voltage and timing for each bath. Applications for tissue bath stimulation include isolated smooth muscle preparations, such as vas deferens, gut and arteries. Other isolated muscle preparations include coronary and skeletal, as well as isolated tension measurements in muscle. The Multistim is said to be a flexible modular stimulator with over 20 components to choose from, allowing the user to configure the system for each experiment. Two case sizes, a 48-cm rack mountable accommodating up to ten stimulation channels and a half-width, bench-top case accommodating up to four channels are available. Several stimulators channel modules are available with voltage protocols up to 100 V and current capacities up to 2.5 A per channel — overload cut-out protection is standard. A choice of timing and triggering modules that allows convenient implementation of pulse train and other stimulus delivery protocols complete the system.

Reader Enquiry No. 114

Spectroscopy

Optiglass

From Starna Brand

Standards extended with Varian-Cary calibration

To complement the extensive range of Starna Brand sealed liquid spectrophotometric standards, Optiglass has extended its range of glass filters to include all those in common use. These filters are suitable



Setting the standards — filters by Optiglass.

for spectrophotometers that are not operated at high resolution. Filters are mounted in black anodized holders to fit standard cell holders, and are enclosed in a velvet-lined case. For wavelength calibration, didymium and holmium filters are available, and can be supplied either singly or as a pair (1GF/D, 1GF/H or 2GF/D/H). Neutral density filters with nominal values of 0.2, 0.6, 1.2 and 1.8 absorbance at A_{540} are used for testing the photometric scale of the instrument. These are normally supplied in a set of four (4ND), but a selection of any two can be supplied (2ND). For evaluation of both wavelength and photometric scales, a full set of six filters is available (6GF). The filters can be supplied with generic data for the wavelengths or absorbance values, or, alternatively, with certification traceable to NIST, with values obtained using a Varian-Cary 4G spectrophotometer.

Reader Enquiry No. 115

S2000 spectrometer

From Ocean Optics

For low-light level applications this is a miniature, fibre-optic ultraviolet/visible-shortwave NIR spectrometer

The S2000 is more sensitive than Ocean Optics' first generation of miniature spectrometers, with a compact optical bench coupled to a 2,048-element CCD linear-array detector. Because of this improved detector technology, the S2000 is better suited for fluorescence, phosphorescence and other low-light level applications. Potential applications include high-resolution chlorophyll analysis, oxygen sensing, plasma monitoring and pharmaceutical testing. The linear CCD array responds to light from 200 nm to 1,000 nm, and can be configured for either ultraviolet/visible/shortwave NIR or visible/shortwave NIR applications. It offers optical resolution ranges from approximately 0.1 nm to 10 nm (FWHM), depending on the groove density, the grating, the core diameter of the input fibre or width of the entrance slit. The system is available as a stand-alone instrument or OEM development component. Users may select either a benchtop or portable spectrometer platform, and choose from among 14 diffraction gratings, plus a variety of input fibres and fibre-optic accessories, to create an application-specific analytical instrument. S2000 spectrometer operating software is available in either DOS or Windows versions. Also, users may add the Ocean32 real-time acquisition, analysis and graphics software, or an application interface package that includes Windows DLL and example programs written in Visual Basic, Borland C, Microsoft Excel and Microsoft C++.

Reader Enquiry No. 116

Biology based

Beacon 2000

From Panvera

A fluorescence polarization system measures equilibrium binding directly in solution

This is a variable-temperature fluorescence polarization (FP) system that measures equilibrium binding and/or molecular degradation for a wide variety of biological molecules. This compact, benchtop instrument measures as little as 10 fmol ml^{-1} of fluorescently labelled sample, allowing the user to determine equilibrium binding constants (K_d) in the low picomolar range. The system provides variable temperature control of the sample from 6 to 65 °C. The instrument's low-volume capability, which requires only 100 μl of sample, makes the system particularly attractive when the amount of protein or other binding partner needs to be conserved. Easy use of different fluorophores is made possible by external filter holders that allow for quick interchange of alternative wavelength filters. It is said to be well-suited to measuring protein-DNA and protein-protein binding, performing sensitive enzyme and drug screening assays, as well as total and sequence-specific DNA quantitation.

Reader Enquiry No. 117

Ultramicro Pipetman and ShaftGard

From Rainin

A system for pipetting microvolumes from 0.1 μl to 10 μl

Ultramicro Pipetman is said to bring improved performance to small-volume measurements with reduced air space for more positive control and a non-protruding piston for protection from contamination and carryover. The FinePoint micro-10 tips are thin-walled, and should provide easier and more consistent droplet removal. The ShaftGard tip is designed to help prevent cross-contamination, with extended length for easy tube access. ShaftGard tips are designed for use with Pipetman P2 and P10 pipettes, as well as PipetPlus LatchMode R2 and R10 and EDP-Plus electronic pipettes. No part of the pipette ever touches the vessel walls so the danger of cross-contamination is significantly reduced. Longer length should provide easier access into microtubes, microcentrifuge and PCR tubes. With ShaftGard tips, the tip ejector is enclosed by the tip and ejects the tip by pressing on the shoulder inside the tip.

Reader Enquiry No. 118

The Megatiter plate

From Continental Laboratory Products

New 96-well, large-volume microplates

The Megatiter plate offers large working volumes in a 96-well plate format. Each plate has 96 2.2-ml wells in the standard 8×12



Megatiter plates with 96 2.2-ml wells.

format. This plate is well-suited for multi-channel pipetting and automated work stations. Made from clarified high-density polypropylene, Megatiter plates are autoclavable and stackable. Each plate has an alphanumeric reference grid for easy sample identification. An adhesive sealing lid is also available to secure the storage of samples.

Reader Enquiry No. 119

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENTS

Gramsch Laboratories
New: Endomorphin -1 & -2 Custom Peptide Synthesis:
 Shipped within one week:
 10 AA DM 400.00; 15 AA DM 520.00;
 20 AA DM 660.00; 10-30 mg unpurified;
 buy only when satisfied with the HPLC profile
Antiserum Production Service:
 you prepare the recombinant protein (in paa-gel or soluble), we produce the antiserum
 Distributor for Japan: FUNAKOSHI CO, Tokio 113
 Gramsch Laboratories
 D-85247 SCHWABHAUSEN, Germany
 Phone +49-8138-1769 or +49-8131-5679-0
 Fax +49-8138-1260 or +49-8131-5679-24

READER ENQUIRY NO. 95

BIABERM®
EQUILIBRIUM DIALYSIS SYSTEM



The gentlest method to quantify the binding parameters and thermodynamic data between ligands and biopolymers. Results come closest to the actual values. Sample volume from 0.2 ml to 5.0 ml. Up to 20 experiments in parallel.

BIABERM
 Gera
 P.O. Box 65012
 D-81215 Munich

READER ENQUIRY NO. 61